

Poor planning threatens centres of excellence

A proposal to convert some of Japan's best basic research institutes into autonomous agencies is fine in principle but could be disastrous without far more thought about the nature of their activities.

Japanese scientists are at last actively entering the debate over the government's controversial reform plan, which targets not only ailing government organizations but also some of Japan's best basic research institutes for conversion to agency status (see page 272). This status would give them independence in management but require them to have their performance evaluated by external assessment bodies.

This publication has long advocated giving greater autonomy and responsibility to the institutions administered by Japan's Ministry of Education, Science, Sports and Culture (Monbusho), including universities and the institutes for joint university use. An independent management system would allow increased flexibility in funding. Furthermore, without the restrictions imposed by the civil service law, researchers would be free to carry out entrepreneurial activities, such as setting up venture businesses and carrying out joint research with private companies.

The government's reform plan moves in this direction but, as it stands, its flaws are far more obvious than its potential benefits. Opposition to reform of the universities has succeeded in delaying further changes until at least 2003, giving time for the fundamental reshaping required. Many fundamental steps need to be taken, such as implementation of a national external review system and of competent new administration, before universities will be ready for agency status. But other institutions would face more immediate problems if the current proposals are implemented.

The reform plan, which is part of the government's drive to improve the country's administration across the board, has met strong opposition from the outset, with targeted institutes arguing against the government's claim that the changes are necessary to run

them more efficiently. Such resistance is hardly surprising given the plan's rationalization targets and cost-cutting measures. The government has also made clear its intention to close down, merge or privatize institutes that fail to meet their performance-related targets. But, following compromises by the government with stronger ministries over its plan to reduce the number of civil servants by 25 per cent, there is disproportionate pressure on research institutes to cut their costs.

As a result, the National Research Institutes for Joint University Use are at risk; they include the Institute of Space and Astronautical Sciences, the National Astronomical Observatory, the National Institute of Genetics and the High Energy Accelerator Research Organization, which are all renowned for the quality of their basic research. More thought must be given to decisions concerning such high-quality institutes, whose activities are ill-suited to targets based on cost performance. The government should create a separate agency plan for institutes carrying out basic research, so that appropriate evaluation systems and performance-related targets can be introduced, with goals and support established over periods of 5–10 years, reflecting the long-term character of fundamental research.

Scientists have been slow to respond to the government's proposals. Partly to blame is the lack of communication between the government and the researchers, but the main reason lies in the fact that the scientific community have very few 'advocates' to represent their opinions. If Japan is to create world-class 'centres of excellence', as stressed in the government's 5-year plan for science and technology, the requirements of basic research institutes must not be overlooked for the sake of a flawed reform plan, while the scientific community needs urgently to develop stronger advocacy. □

Don't ignore dual careers

The dangers of doing so are all too vivid in the physics community.

Why are only 6 per cent of US physicists female? And why is women's progress into the upper echelons of the hard sciences so painfully slow? A survey due to be presented this week at the American Physical Society's annual meeting reveals that being part of a dual-career couple can form a critical barrier to the progression of the small number of women working in physics (see page 273).

Many more women physicists than men are married to other physicists, so are in the complicated position of having to search for a job with a physicist spouse. Worryingly, the survey finds that dual-career couples working in science are put in a worse situation by what is described as "institutional ignorance", "bureaucratic inertia" and what can be summed up as plainly sexist ideas about the role of women in dual-career couples. We should not be overly surprised at this — witness, for example, a recent report from the Massachusetts Institute of Technology that women

faculty are systematically rewarded less for professional accomplishments than their male colleagues.

The problems faced by dual-career couples — in any profession — pose a challenge for institutions. If the small numbers of female physicists are to be retained, their employers need to stop ignoring the issue or even making things worse. Most institutions have been slow to identify and deal with the obstacles to employing dual-career couples, yet there are a number of tried and tested ways they could adopt — for example, offering shared or split positions, or running spousal hiring programmes. As the survey shows, these ideas have already been successfully put into practice at a few US higher-education institutions.

With increasing numbers of women taking up scientific careers, the number of dual-career couples in science is likely to continue to grow. Institutions have no option but to face that phenomenon and adapt. □



Nuclear waste store could be built within 25 years, say Lords

[LONDON] The British government should immediately resume planning for disposal of civilian and military nuclear waste underground, according to a report published this week by a House of Lords committee.

The Lords Select Committee on Science and Technology suggests that such a repository could be planned and built within 25 years. The findings come two years after environmentalists and the local authority managed to kill the UK government's plans to build a permanent store at Sellafield, Cumbria (see *Nature* 386, 423–424; 1997). They will re-ignite a controversy over the disposal of more than 70,000 cubic metres of nuclear waste in Britain that is currently stored above ground.

The report acknowledges that, although there is an international consensus in favour of deep underground disposal of radioactive wastes, this is not always a popular option. It says the government needs to build public trust in the concept by starting public consultation on a new national radioactive waste management policy. A central aim of the policy should be to give the public and parliament more say in decision-making than they have had in the past.

Local authorities, the report says, may need incentives to "volunteer" to host potential repositories, as happens in other countries. It adds that residents near a potential repository site may need compensation for the adverse impact on house prices, businesses and other forms of "blight" experienced by communities with nuclear neighbours.

The committee says that any new policy should also cover wastes from the nuclear



Dig deep: nuclear waste stored on the surface at Sellafield could yet go underground.

weapons programme, and plutonium from defence sources and civil nuclear power plants. Until now, plutonium has not been classed as a waste owing to its potential use as a reactor fuel. The report says that "there is no reason to store plutonium which is surplus to all foreseeable requirements".

A repository should remain open for research and monitoring until the government is satisfied that the wastes pose a negligible threat to human health and the environment, says the report. The repository would then be sealed off, although retrieval of wastes should still be possible.

A key recommendation comes in the form of a blueprint for a new policy structure for managing radioactive waste. The committee calls for the establishment of a new statutory body, the Nuclear Waste Management Commission. This should draw up a

list of possible sites, commission and monitor research, and oversee construction of the repository. The commission's members would be drawn from a wide range of backgrounds, and it would report to parliament.

A Radioactive Waste Disposal Company funded by the nuclear industry would carry out site investigation, research and repository construction. Its work programme would need approval from the commission, and its research would be published and peer reviewed. The company would decide the final choice of site, but this would need to be ratified by both houses of parliament before being put to a public inquiry.

The structure of the two bodies is similar to that envisaged more than 20 years ago in a report from the Royal Commission on Environmental Pollution, which was chaired by Brian Flowers, former vice-chancellor of the University of London, who is also a member of the Lords committee.

Successive governments only partially implemented the Flowers recommendations. The government's existing Radioactive Waste Management Advisory Committee, for example, was given no powers to commission research, and has no representation from environmentalist groups.

Nirex, the nuclear-industry-funded company set up in 1982 to manage waste from civilian nuclear power plants, suffered from a reputation as an excessively secretive organization, beholden to the nuclear industry and reluctant to publish its findings.

The Lords report says that no solution will be viable without public acceptance. But, like the government, the Lords committee is unconvinced that the views of environmentalist groups are the same as those of the broader public, and is keen to explore ways of establishing those views.

"Some new methods are currently being tried which do not give undue preference to minorities at the expense of the 'silent majority'," the report says. This is partly a reference to an ongoing consensus conference, in which a panel of lay people will in the coming weeks deliver its verdict on the future of radioactive waste.

Environmental groups such as Greenpeace insist that radioactive wastes must be stored above ground permanently on or near sites where they are produced, both to reduce transport journeys, and so that the wastes can be monitored and retrieved if necessary.

Other groups want the wastes to be stored above ground at least while research and development on the safety of the deep repository concept is completed. **Ehsan Masood**

Gaps remain in knowledge of waste behaviour

Although deep disposal is its preferred way to deal with radioactive waste, the House of Lords Select Committee on Science and Technology acknowledges that many technical questions about this remain unanswered.

"Initial 'guesstimates' about the rates at which waste packages will corrode and the rates at which radionuclides will be leached out into groundwater have been replaced by firmer estimates based on a firmer understanding of the physical and chemical processes

involved," says the report.

But it adds that many gaps in the knowledge remain. These include the effects of earthquakes and future changes in climate on the rates and patterns of groundwater flow. The report says that movement of groundwater through ordinary fractured rock is difficult to model, without factoring in the effect of an earthquake.

The report adds that, although changes in sea levels and the timing of future glaciations can be predicted, the means to

predict the effects of climate changes on groundwater flow are incomplete. It adds:

"Techniques such as palaeohydrogeology, in which attempts are made to reconstruct past groundwater conditions as an indicator of future ones, are in their infancy."

A "further major type of uncertainty", it says, is the extrapolation of the results of short-term laboratory experiments to thousands of years and longer. This applies to canister corrosion and waste leaching. **E. M.**

Unease mounts over Japan's reform plan

[TOKYO] Japanese scientists are increasingly concerned about a government plan to turn 14 basic research institutes attached to the Ministry of Education, Science, Sports and Culture (Monbusho) into semi-autonomous 'agencies' with greater administrative independence.

The National Research Institutes for Joint University Use — which include the High Energy Accelerator Research Organization, the National Institute of Genetics (NIG) and the Institute of Space and Astronautical Sciences (ISAS) — are the latest target of the government's planned administrative reforms (see *Nature* 398, 5; 1999).

These are stand-alone entities with their own research staff, facilities and research agendas, but they are also used by university researchers, and some institute staff have affiliated positions at universities.

The plan, based on the UK 'Next Steps' initiative, seeks to improve administration by restructuring government ministries and agencies. It would create semi-autonomous bodies with separate management systems. Each institute would have its performance evaluated by an external assessment body every 3 to 5 years.

The plan initially included research institutes attached to science-related ministries, such as the National Institute of Radiological Sciences (see *Nature* 395, 211; 1998). But the government has recently made a proposal to turn institutes for joint university use into semi-autonomous institutions as well.

The government hopes to reduce the number of civil servants by 25 per cent within the next decade by giving agency status to government-run organizations. But critics argue that the government's aim is ambiguous, as they have announced that most of the agencies' staff will remain as civil servants.

Junichi Tomizawa, former director of NIG, says he is afraid that the restructuring will be used "as a scapegoat to cover up a possible failure to reduce the number of civil servants at other organizations".

Although many researchers support the plan's basic principles, such as giving the institutes greater flexibility in funding, they are opposed to its emphasis on performance-related targets, which they say are inappropriate for basic research laboratories.

According to policy documents released in January, agencies would have an accounting system like that of private companies, with performance targets related to costs. But the government has now revised these targets, so that financial performance would not dictate the priorities of the research institutes.

But researchers are still concerned that the targets the research institutes would have to meet have not been adequately specified by the government. "Although the emphasis on cost performance is not necessarily explicit in the policy outlines, we are worried that the plan has been made within the framework of the administrative reform plan, which is aimed at reducing government costs" says Yoshiaki Hotta, director of NIG.

Hotta also suggests that the performance of research institutes should not be judged in periods of 3 to 5 years, given the long-term nature of basic research. "Excellent scientific contributions are not necessarily regarded as good during their initial stages," he says.

Some researchers are calling for a separate agency plan for basic research institutes, so that appropriate targets and evaluation systems can be introduced.

Tomizawa says the needs of the institutes and the national universities should be handled jointly. He is worried that their close relationship would be jeopardized if the institutes fall under a different administrative system to that of universities.

The government would like to turn universities into agencies, but has had to abandon this in the face of strong opposition from the academic community (see *Nature* 395, 730; 1998). Akito Arima, the education minister and director-general of the Science and Technology Agency, has declared that such reforms would harm both education and research. The government says it will not make a final decision on how the reforms will affect the national universities until 2003.

Minoru Oda, former director of ISAS, warns the universities against complacency. "Some of the powerful imperial universities, such as Tokyo and Kyoto, have this sense of security that they will not be subjected to any radical changes, while others have simply given up, thinking there is no way out from this agency plan," he says. **Asako Saegusa**

Libraries offer incentive for web-based rivals to 'costly' journals

[WASHINGTON] A coalition of university libraries is stepping up its battle to circumvent the high price of specialized scientific journals by offering \$0.5 million dollars to universities and professional societies to come up with web-based information resources at universities.

The Scholarly Publishing and Academic Resources Coalition (SPARC), an offshoot of the Association of Research Libraries (see *Nature* 393, 719; 1998 and *Nature* 397, 195; 1999), last year helped to launch several journals to compete with publications which the libraries consider too expensive.

The coalition, which has 162 members, will make awards totalling \$500,000 to developers of web-based venues for publishing research. The money was raised from the \$5,000 annual dues of SPARC members, including the libraries of the universities of Harvard, Princeton, Johns Hopkins and California, as well as Canadian, Danish, Australian and British universities.

The idea, says Rick Johnson, SPARC's executive director, is to get applicants to find

"new ways of disseminating research that are really responsive to the needs of scientists" and are cheaper than journals.

The aim is to create electronic repositories where research can be posted, discussed and seen by scientists worldwide, without waiting for journal publication.

The best-known model of such a resource — the Los Alamos e-print archives — is a "macro" example of what SPARC hopes to fund, says Jim Neal, dean of university libraries at Johns Hopkins University and a member of the SPARC steering committee. But he says proposals may range in strategy and structure.

One application for the SPARC awards is likely to come from Columbia University, where an earth-sciences web resource is being developed. *Columbia Earthscape*, a web-based publication scheduled for launch this autumn, will publish papers in their formative stages, early versions of conference proceedings, data sets and a magazine on science policy.

"We're not just trying to mimic on-line

something that's available in print," says Kate Wittenberg, editor-in-chief at Columbia University Press. "We are trying to figure out ways to make very innovative, cutting-edge materials available at a cost possible for libraries to afford."

One weakness is how to entice scientists to use such sites when they lack the prestige of peer-reviewed, print journals. This is why support for the proposals from scientific societies is vital, says Johnson. "The prestige that a society can breathe into a venture is going to be very important."

In July, SPARC will help to launch a third journal to compete with an established publication. *Organic Letters*, to be published by the American Chemical Society, will be available in both print and electronic forms for \$2,300. Its competitor, *Tetrahedron Letters*, published by Elsevier, costs \$9,862.

SPARC supports alternative journals by delivering its library members as customers. So far, it has helped to launch *Evolutionary Ecology Research* and *PhysChemComm*, an electronic journal. **Meredith Wadman**

Husbands a drag to high-flying physicists

[LONDON] Many women physicists in the United States face special obstacles in their career progression because they are married to physicists or other scientists, according to a new study.

Nearly 45 per cent of married female physicists have physicist spouses, though only 6 per cent of male physicists are in the same position. This means that many female physicists looking for a job have a husband in the same field — where, typically, few jobs are available, say Laurie McNeil of the University of North Carolina and Marc Sher of the College of William and Mary in Virginia. They will present their findings to a meeting of the American Physical Society (APS) in Atlanta, Georgia, this week.

Compounding the problem, university departments tend to be unresponsive to dual-career issues, according to the study, which was undertaken for the APS's Committee on the Status of Women in Physics.

The authors say that the problems they have identified represent significant constraints on the growth of the small community of female physicists.

In a survey of 620 dual-career couples where one partner was a physicist, which McNeil and Sher made last year, 60 per cent said that one partner had taken a lower-level science job, a non-science job or no job at all owing to the difficulty they had in finding two scientific appointments at the same location.

Although some institutions have tried to find solutions to the problem — such as creating split or shared positions or spouse-hiring programmes — most respondents said that, in their experience, institutions tend either to handle dual-career issues badly or to ignore them altogether.

Members of couples who were both pursuing careers in science said they had been given less consideration for a post, or none at all, when it became clear they were part of such a couple. Rules against nepotism are often, incorrectly, cited as a reason why an institution is unable to help spouses find employment there.

"There are ways to address these problems — and it's very important for institutions to address them, as they won't go away," says McNeil.

Judy Franz, the executive officer of APS, says: "Universities rarely deal with issues until they have to. In physics, until recently, the majority of departments had no women faculty and therefore never had to deal with this challenge."

The report points out that many institutions view dual-career couples as an administrative problem, rather than an opportunity, and are unable to see the benefits that dual-career couples can bring. McNeil and Sher say that institutions



offering options for dual-career couples will, in fact, attract the best staff and will stand a better chance of retaining them.

The University of Illinois, for example, has established a dual-career couple programme, which focuses on improving recruitment and retention of faculty members when the appointment or retention of one partner is contingent upon the employment of another.

Franz was taken aback by the lack of progress that the survey reveals, especially

because some institutions have had policies in place since the 1970s.

"I was most surprised by the negative attitudes expressed," she says. "I thought that they had mostly gone by now."

One potential employer told a candidate that his partner "shouldn't be working anyway". Another candidate was told that his wife should make herself available to do voluntary scientific work, since it was his role to support the family.

"A lot of remarks and responses sounded like something of an earlier era," says McNeil. "You wouldn't think people behaved in that way, although you might have expected it 20 years ago."

With greater numbers of women entering physics, institutions hiring physicists are increasingly likely to find that their best candidate has a spouse seeking professional employment. Institutions have a role to play if the number of female physicists is to increase, McNeil and Sher say. "It is within the power of the institutions to help ease the situation or to make it worse," their report says.

The full 60-page report, with relevant links, is available at: <http://www.physics.wm.edu/dualcareer.html>

Natasha Loder

California approves \$1.5 billion campus

[SAN DIEGO] The University of California at San Francisco (UCSF) has begun developing a new life-sciences research campus at a cost of up to \$1.5 billion.

The University of California Board of Regents last week approved the first building of the new Mission Bay campus. The \$222 million complex will house centres for structural and chemical biology and for molecular and cell biology.

Plans for a second building, with laboratories for genetics, developmental biology and neurobiology, at a cost of about \$100 million, are near completion. These plans are expected to be approved soon.

The Mission Bay campus is designed to replace much of the present UCSF research facilities adjacent to its medical centre at the Parnassus Heights campus, where buildings have been damaged by earthquakes, laboratory space is at a premium, and expansion is limited by residents' objections.

The 43-acre Mission Bay campus, to be built on the site of a former railway yard and warehouse district in the south-eastern part of the city, will be ringed by private development of biotechnology and related firms that hope to capitalize on the new research complex.

The project is being compared by its supporters to the Stanford Industrial Park,

which formed the nucleus of Silicon Valley near Stanford University.

"This represents a dramatic opportunity for our campus," said Zach Hall, a neuroscientist who is UCSF's vice-chancellor for research. "Most of our existing basic research faculty will move to the new campus and there will be a significant number of new faculty and associates in the future."

Hall could not specify how many new research positions will be created, but said that as lab space grows, so too will staff and educational opportunities. The first building will have laboratories for about 50 principal investigators and will eventually house 900 personnel. When completed, the Mission Bay campus will provide work space for about 9,000 scientists, graduate students and staff in 20 buildings.

Current plans call for focusing on basic research at the Mission Bay campus, with disease-related research efforts remaining at the Parnassus Heights campus near the UCSF medical centre.

The cost of the first building on the Mission Bay campus is being met by \$110 million in gifts, \$70 million in bonds, \$21 million in state funds, and \$18 million in lease payments from the Howard Hughes Medical Institute, which will occupy part of the building.

Rex Dalton

Dispute erupts over Nazi research claims

[MUNICH] German science suffered less damage during the Nazi era than has been commonly assumed, despite the forced exodus of Jewish researchers and some of the worst medical crimes ever reported, according to a controversial new book.

Author Notker Hammerstein, a respected professor of history at the University of Frankfurt, argues that only a few non-Jewish scientists felt threatened by the Nazis, despite the regime's anti-intellectualism. His book on the history of Germany's research council, the Deutsche Forschungsgemeinschaft (DFG), *The DFG under the Weimar Republic and the Third Reich*, was commissioned by the DFG.

Although all Jewish scientists were expelled from their university chairs as one of the earliest acts of the Nazi regime, the quality of research did not fall significantly behind other countries, Hammerstein claims.

He argues that the academic mainstream in Germany adjusted rapidly to the new political environment. Academics were mainly concerned to save or promote their own careers and disciplines. They failed to recognize that 'normality' served an illegitimate totalitarian system, he concludes.

In 1933 the DFG was forced to abandon its founding principles, and peer review was replaced by favouritism and arbitrary decisions, says Hammerstein. When Rudolf Mentzel, a 36-year-old chemist, was made president in 1936, "the DFG basically stopped existing".

But, given the extraordinary circumstances, Hammerstein finds that a relatively normal academic life was maintained at universities. Although Hitler and the



Rudolf Mentzel, whose presidency effectively wiped out Germany's research council.

Wehrmacht bosses had little interest in research, good science continued to get money. "Mentzel could not afford to fund only dead-losses and ideologists," says Hammerstein.

The ideological blindness of some scientists was extreme. Physicists such as Mentzel's predecessor Johannes Stark, and Philipp Lenard — both Nobel prizewinners and convinced anti-Semites — proclaimed a vaguely romantic "German Physics", which denied Einstein's "Jewish" theory of relativity.

Hammerstein's assessment has already been disputed. "The emigration of Jewish scientists led to a crucial decrease of Germany's intellectual competitiveness from which we still have not recovered complete-

ly," says Ernst-Otto Fischer, who won the Nobel prize for chemistry in 1970.

Ernst Klee, author of several books about medicine in the Nazi era, says that Hammerstein has downplayed the DFG's involvement in medical war crimes. After 1940, funding applications were approved for research that included lethal experiments on humans. The DFG also supported what it called research into "the genetic disposition of the offspring of vagabonds, robbers and gypsies", but which historians believe was nothing more than a front for their extermination.

Hammerstein says that most DFG funds supported 'normal' research, unrelated to Nazi ideology, but Klee accuses him of simply trying to save the DFG embarrassment.

The book was commissioned in 1995 by former DFG president Wolfgang Frühwald to mark the research council's 75th anniversary. Frühwald said at the time that a critical assessment would be more appropriate than a "normal sort of celebration".

Wolfgang Schieder, a professor of history at the University of Cologne and an expert on fascism, defends Hammerstein. Sensitivity to Nazi issues, he says, often causes emotional reactions: "It is much easier to write about an eighteenth century problem."

The Max Planck Society (MPS), Germany's main non-university research organization, is also confronting its Nazi past. A five-year research programme by non-MPS scientists will examine the history of its pre-1945 predecessor, the Kaiser Wilhelm Society. It was launched this month with an international meeting in Berlin on science during the Nazi era.

Quirin Schiermeier

Embattled French science agency says it is holding its own

[PARIS] The Centre National de la Recherche Scientifique (CNRS) is by far the largest player in French research and its output is internationally competitive, it claims in a new study.

The French science agency is aiming to answer its critics: for over a year, it has been on the defensive following repeated attacks by the science minister Claude Allègre, whose reform plans would strengthen the universities — to the detriment of CNRS.

The agency is now becoming more assertive, bolstered by the recent ministry retreat on plans to impose reforms on it. Following a rebellion in the scientific community, the ministry ceded to CNRS demands for a national debate and a rethink of the reform plans in concert with the agency (see *Nature* 397, 463; 1999).

The study assesses the output and impact of papers with at least one CNRS author from 1986 to 1996, using journals indexed

by ISI in the United States. It found that CNRS contributed to 52 per cent of the 370,000 French publications — those with at least one French address — published over the period in all 'hard' scientific disciplines.

This share rose to around 80 per cent in journals of physics, chemistry, astronomy and astrophysics. CNRS's share of French papers in biomedical journals was less than 20 per cent, but biomedical research is specifically catered for by another French agency, Inserm.

Papers with a CNRS author also had more impact than those from other French laboratories, according to the study. Its authors calculate that CNRS accounts for about 80 per cent of French impact in physics, chemistry, astronomy and astrophysics, two thirds in basic biology and 28 per cent in biomedical research.

At the international level, the average impact (in the two years following

publication) of CNRS papers across all disciplines was — at around 1.2 — equivalent to that of papers produced by all US scientists in ISI's Science Citation Index. CNRS papers made more impact than French papers overall, and more than those by German, British and Japanese authors.

Jacques Sevin, director of strategy and programmes at CNRS, says that the results are not a cause for complacency and improvements are still needed. But he argues that the study's conclusions provide an unequivocal response to outside pressure on CNRS to justify its performance.

France's share of the world's publications also rose from around 5 per cent to 7 per cent over the period surveyed. This is a respectable performance, say officials, given that the 'market share' of all nations is under pressure owing to sharp increases in the number of publications from emerging economies.

Declan Butler

CERN plan to train science entrepreneurs

[GENEVA] CERN, the European Laboratory for Particle Physics, is stepping up its technology-transfer activities in response to pressure from its 19 member states.

The Geneva-based laboratory has created a directorship for technology transfer to oversee an expansion of its training programmes for young scientists and engineers and of its patenting activities.

The changes are being implemented by CERN's director-general Luciano Maiani, former president of the Italian particle physics organization INFN, who joined CERN last year. The INFN also has an action plan of its own to promote technology transfer.

CERN has appointed a director of technology transfer, Hans Hoffmann, who will take up the position in June. The laboratory puts 1,000 young scientists each year through its doctoral and postdoctoral training schemes, and Hoffmann estimates that more than half of them move immediately into industry, mainly to take up jobs as information scientists.

The laboratory will now offer these people additional courses in entrepreneurship and intellectual property rights, to encour-



Industrial awareness: CERN training will include business studies and intellectual property rights.

age them to start their own companies when they finish training.

It will also keep a sharper eye out for patentable inventions. CERN has been criticized in the past for failing to profit from important inventions such as the World-

Wide Web, which was developed in 1989 by Tim Berners-Lee during his six-year stay at the laboratory.

The laboratory has filed only a handful of patent applications in its entire history, says Hoffmann. It has operated on the basis that, as a basic research organization, all information and discoveries should be made freely and immediately available. Hoffmann says this will continue to be the case — but that the publication of some information may be delayed until decisions have been made about patent filing.

CERN is also sensitive to concerns of its member states that the licensing of patents should be fairly distributed between them. Some members, including the United Kingdom, have complained that patents taken out by former CERN director general Carlo Rubbia were transferred to a company in Spain, without allowing other countries to put forward any interests of their own. The patents related to technologies involved in Rubbia's so-called energy amplifier, which uses a particle accelerator to drive a subcritical reactor which burns highly radioactive nuclear waste to generate energy.

INFN, meanwhile, hopes to use European Union structural funds — subsidies to poorer regions — to support an industrial training scheme which it plans to launch in the summer.

The scheme, which the organization is currently discussing with industry, would employ large numbers of industrial trainees for short periods in INFN institutes in the south of Italy.

INFN president Enzo Iarocci comments: "We will maintain our identity as a basic research organization, while responding to pressure from industry to 'give something back'."

Alison Abbott

'Don't try to change embryo research law'

[MUNICH] Now is not the right time to campaign for a relaxation of Germany's tight laws on human embryo research, as some scientists have demanded, according to a policy statement issued last week by the Deutsche Forschungsgemeinschaft (DFG), Germany's basic research funding agency.

Any attempt to reconsider the law could backfire, explains Rüdiger Wolfrum, vice-president of the DFG, because the public is not fully aware of the medical potential of human embryo research.

The DFG has assessed the existing laws to establish exactly how German researchers are placed to conduct research related to embryos. It finds that they are free to use one particular source of human embryonic stem cells — which are still capable of dividing and developing — for the many different lines of medical research which the DFG considers justified, including the development of cell transplantation therapies for incurable diseases such as Alzheimer's and Parkinson's.

The law does not allow researchers to use embryos developed from spare eggs collected during *in vitro* fertilization. The Embryo Protection Law confers a human egg with full human rights from the moment of fertilization. Attempts to bypass fertilization by transferring genetic material directly into an enucleated human egg —

cloning — are also expressly forbidden.

But there is no legal impediment to scientists isolating primordial germ cells from aborted fetuses and culturing them under conditions that allow them to develop into pluripotent human stem cells, which are committed to developing into one particular sort of body tissue. Under normal conditions, a fertilized egg divides initially into totipotent cells, which have the potential to develop into any sort of tissue, and thus theoretically into a complete human, but they become pluripotent after a few cell divisions.

Scientists have already established that pluripotent stem cells may be generated from fetal material in mice. Studies in the United States indicate that this principle will also apply to human material (see *Nature* 397, 185; 1999), allowing researchers to bypass the need to handle legally protected totipotent cells. The DFG says it will support studies that further this area of research.

The DFG is aware of the ethical sensitivity of the issue. It calls for the establishment of a central committee to assess the ethical, legal and scientific implications of research proposals involving the use of human embryonic stem cells, and to monitor any projects undertaken. It also calls for a full and open discussion of all these issues at European level.

A. A.

Weapon lab security may hit foreign visits

[WASHINGTON] Security is being tightened at the US Department of Energy's major weapons laboratories in the wake of the dismissal of a Los Alamos National Laboratory engineer for alleged involvement in a security breach.

But scientists are worried that the political reaction to the leak, which may have occurred more than a decade ago, will reduce or end contacts between the laboratories and foreign scientists (see *Nature* 398, 96; 1999).

Wen Ho Lee was fired from Los Alamos not for espionage, which remains unproven, but for talking to Chinese scientists and failing adequately to report the conversations. Lee, a 20-year veteran at Los Alamos, had been working under contract for the laboratory. The energy department believes he may have aided Chinese visitors in gathering information on the H-88 missile warhead.

The incident at Los Alamos has touched off a political storm in Washington. Republicans in Congress are eager to use the issue to score political points against President Bill Clinton and Vice-President Al Gore, who will be running to succeed Clinton in the 2000 election.



Richardson: keen to protect visiting scientists programme.

John Pike, a defence analyst at the Federation of American Scientists in Washington, says: "They are making this sound like Los Alamos is a [convenience store] with H-bombs ready to go. The real question is whether Lee walked out the front door with the H-88 warhead under his arm or whether he ran off a little at the mouth in conversation over dinner."

The General Accounting Office has already published several studies criticizing security procedures at the three weapons laboratories, Los Alamos and Sandia in New Mexico and Lawrence Livermore in California.

John Browne, the director of Los Alamos, insists that the security breach was real and the firing was deserved. Lee was not fired for theft of nuclear secrets, says Browne. "He lost his job for security violations, not for espionage."

Last Friday (19 March) a delegation from Washington visited Los Alamos to review the new security procedures. They included Ernest Moniz, the energy under-secretary, Vic Reis, the assistant secretary with responsibility for nuclear weapons, and Ed Curran, head of the energy department's Office of Counterintelligence.

The initiatives include an additional \$8 million on the department's 2000 budget request to start a 'cyber-information' security programme. The amount would raise the department's proposed counter-intelligence budget to \$39.2 million.

Other steps include measures to control weapon design data in secret documents at the laboratories; the appointment of former Central Intelligence Agency director John Deutch to review the energy department's foreign visitors and assignments programme; and a review of how the department maintains its counter-intelligence files.

Extra 'counter-intelligence professionals' have been appointed to the laboratories, and procedures for screening foreign visitors have been tightened. The Office of Counter-

intelligence, which was buried within the agency's bureaucracy, has also been raised in status, with Curran reporting direct to Bill Richardson, the energy secretary.

One question now arising is how many laboratory scientists, apart from Lee, ought to be dismissed for lax security practices, such as copying classified e-mails or failing to fully report conversations with foreigners.

Another is whether the Lee incident will lead to the suspension of visits of scientists from sensitive countries, such as China. Los Alamos says 100 Chinese nationals worked at the lab last year and 278 others paid visits.

Richardson has expressed concern about the political reaction to the incident. "I'm concerned Congress is going to over-react. You don't eliminate the foreign visitors programme because it is essential." **Wil Lepkowski**

Trade unions campaign for lifelong learning

[LONDON] Britain's largest science trade unions have called on scientists and employers to pay closer attention to continuing professional development — or lifelong learning — to ensure that scientists remain at the forefront of their profession.

The unions unveiled a 14-point charter on training and development for scientists at a conference in London last week. The initiative was endorsed by the Association of University Teachers, the Institute of Professionals, Managers and Specialists (IPMS), the Manufacturing Science Finance union, and the National Association of Teachers in Further and Higher Education.

The charter calls on employers to set aside at least 2 per cent of their budgets for training and development of scientific staff. It says that scientists should be allowed to take at least 10 days off each year for training and development. And it says that employers should publish performance indicators on their training schemes in their annual reports.

"Those working in universities would need far more than this to keep up with relevant literature and attend conferences. But, all too often, teaching pressures squeeze out time for professional development," says the charter.

Baroness Margaret Sharp, of the Science Policy Research Unit at the University of Sussex, commented: "Most of my staff are working 60 to 70 hours per week [trying to meet] targets for teaching quality and the research assessment exercise. They are too exhausted to train. There just isn't enough time."

The conference heard that Britain's army

of contract researchers, technical staff, and scientists from minority communities were often most in need — but also most likely to miss out on — professional training and development.

In the case of scientists, this is partly because of lack of opportunities, and partly through ignorance of the facilities available, the conference was told. Technical and support staff, on the other hand, miss out on training and development because of a lack of motivation, according to Sue Ferns, a research officer at the IPMS.

Ferns added that scientists often benefit from training in areas outside their scientific expertise, for example in administration or finance, which helps if they need to switch careers. The IPMS has established training partnerships with employers. One scheme enables staff to take a Master of Business Administration degree in technology management from the Open University.

One speaker said that employers are reluctant to fund training for contract staff because they "see themselves as consumers of trained staff, not trainers of people".

Other speakers told the conference that employers are sometimes afraid to invest in people because of a fear of losing their investment if a trained employee leaves the company.

But Sir Geoffrey Allen, chairman of the Science, Technology and Mathematics Council, and a former head of research at the Unilever company, said many large companies expect newly trained, but long-serving, staff to move on. It is an opportunity for the companies to refresh employee teams, he said. **Ehsan Masood**

Minister slams 'mismanaged' universities

[CAPE TOWN] Sibusiso Bengu, South Africa's outgoing Minister of Education, has launched a stinging attack on the mismanagement of the universities that were set up for non-white ethnic groups during the apartheid era.

Bengu accused the universities' administrators of "invoking history to hide patent mismanagement". Bengu, who will retire from the cabinet after the June election, launched his attack in an article in the country's financial newspaper, *Business Day*.

His comments, coming from a black minister, will increase pressure on the historically black universities, which have been losing able black students and staff to top universities that were closed to them under the apartheid system of government, which collapsed in 1994.

The historically black universities are a legacy of apartheid, which created separate universities for each major ethnic group, including the mixed-race (coloured) population and the Indian population, while restricting black enrolment at the Afrikaans and English universities.

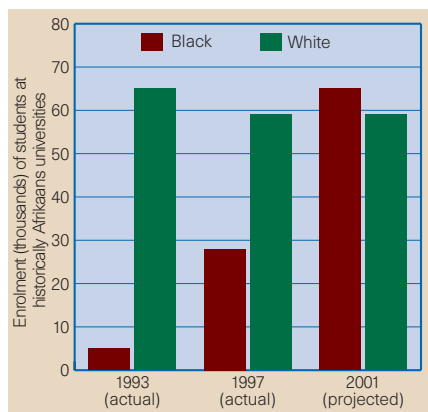
Since the new government took office, the historically disadvantaged universities (HDUs), as they are known, have benefited from special "redress" funds from the state, in addition to normal operating funds. But Bengu is now claiming that "their fixation with historical disadvantage is blinding them to the real challenges they face as universities".

Ironically, Bengu singled out for criticism the University of Fort Hare in the Eastern Cape Province town of Alice, where he was rector until he entered government. Between 1993 and 1997, the government more than doubled its support for the university. But its finances showed a surplus of R11 million (US\$1.8 million) at the beginning of the period and a deficit of R24 million at the end.

Bengu's article followed the release of a report by Stuart Saunders, former vice-chancellor of the University of Cape Town, recommending that the government give serious consideration to closing down the University of Fort Hare. This month, Fort Hare rector Mbulelo Mzamane called for the establishment of a "Marshall plan" to assist the HDUs. It seems unlikely that the government would close the university, which is the alma mater of President Nelson Mandela and of the Minister of Home Affairs, Mangosuthu Buthelezi, leader of the Inkatha Freedom Party.

But last week Bengu appointed private auditors to investigate the books of Fort Hare and five other HDUs: the universities of Zululand, the North, the North West, Transkei, and the Medical University of South Africa.

Last Friday (19 March), students and staff at Fort Hare took part in a protest demand-



Voting with their feet: black students are flocking to historically Afrikaans universities.

ing the dismissal of Mzamane and two senior managers. Saunders' report alleged that the rector was using university funds to pay 75 per cent of his daughter's fees at Boston University in the United States, and recommended that Mzamane's contract should not be renewed.

Cecil Abrahams, rector of the University of the Western Cape, said he was "surprised" by the minister's statement. "I didn't think we needed to educate the minister of the issue of disadvantage," he said, adding that he not only supported the minister's action in setting up the audit, but felt that it should be extended to all the country's universities. "Only then will we be in a position to assess which institutions are viable," he said.

The woes of the HDUs can partly be accounted for by the non-payment of fees. It

is believed that South Africa's institutions of higher education are owed more than R700 million in outstanding fees. The government has set up a fund to lend fees to poor students, but, according to Abrahams, it is able to support only a third of those who are eligible.

Ironically, the gap between the HDUs and the historically Afrikaans universities has widened since the new government took office. Nico Cloete, of the Centre for Higher Education Transformation, says that consequently there is a shake-up in higher education, with "good quality black students moving to the institutions of their choice".

The historically Afrikaans universities have experienced rapid growth in enrolment of black students. Between 1993 and 1997, black (including African, coloured and Indian) student registration at these six universities quadrupled from 9 to 36 per cent.

In the case of black Africans, the most disadvantaged sector of the population, there were 26,000 students registered at the historically Afrikaans universities in 1997, compared to only 16,000 at the four traditionally English universities.

The drop in registrations at the HDUs can be also be attributed to increased numbers preferring to study at technical colleges offering more market-orientated qualifications.

Funding agencies may be forced to rethink their strategies for targeting disadvantaged students. South Africa's Foundation for Research Development and Britain's Royal Society are among those that have earmarked special funds for developing research capacity at HDUs. **Michael Cherry**

UK life scientists seek wider Foresight role

[LONDON] Britain's life scientists are lobbying to be allowed to retain a formal role in the government's Foresight exercise following a decision by the government last year to abolish the Foresight life-sciences panel.

Some life scientists remain critical of the decision to abolish the panel, and many are concerned at the diminished role played by scientists in the new round of Foresight.

Last week, representatives of seven professional societies, including the UK Life Sciences Committee, the Association of the British Pharmaceutical Industry, and the Academy of Medical Sciences, met in London to resurrect the life-sciences panel under the guise of an 'associate' panel, which is allowed in the new rules. One society representative who attended the meeting says the panel is being reconstituted to give life scientists an opportunity to contribute to the new Foresight process.

A spokesman for the government's Office of Science and Technology confirmed that a

preliminary approach has been made by life scientists to set up an associate panel, although no decision has been made about whether permission will be given.

Technology Foresight, which was established by a Conservative government in 1993, asked scientists and industrialists in 16 areas of the economy to determine how research could help to create wealth and improve the national quality of life.

In its reorganization of Foresight last year, the Labour government broadened the exercise by inviting social scientists, natural scientists and industrialists to focus on the government's social and environmental priorities, including ageing, climate change, crime control, and sustainable development. But certain specialist panels were abolished.

The life sciences panel was replaced by a broader 'healthcare' panel, whose chairman is expected to be Michael Peckham, the head of social policy at University College London. **Ehsan Masood**

One third of population will go thirsty by 2025, says UN report

[WASHINGTON] Almost one third of the world's population is expected to face water shortages by 2025, an increase of 10 per cent on present numbers, according to an analysis released last week by the United Nations Environment Programme (UNEP).

The demand for water is increasing at twice the rate of global population growth, largely because of rising living standards and increased food production, according to Hans van Ginkel, rector of the UN University. UNEP says that providing clean water and sanitation to those in need would require a trebling to US\$24 billion of the current global annual investment.

The report says that, in many countries, water shortages arise from inefficient use, pollution and the unsustainable use of underground water in aquifers. For example, 40 to 60 per cent of water used by utilities is lost to leaks, theft and poor accounting.

Tough law helps cut lab animal population

[MUNICH] The number of animals used for research in Germany fell by 40 per cent

from 1991 to 1997, to a total of 1.5 million, according to a federal report published last week. The report attributes the decline to Germany's strict animal protection law, and to the development of alternatives such as computer models.

Mice, rats and guinea pigs formed 80 per cent of the animals used. The report also records 1,900 experiments with primates. Almost half the animals were used for pharmaceutical research.

The report does not reveal the number of research animals that died.

German state enforces code to combat fraud

[MUNICH] Baden-Württemberg has become the first German *Land* (state) to make funding of university research conditional on a binding commitment to good scientific practice.

Under a proposed amendment to the state university law, institutions will be required to follow guidelines on good scientific practice published in 1997 by the Deutsche Forschungsgemeinschaft, Germany's main grants agency.

The good practice guidelines were drawn up following a major fraud scandal in Baden-Württemberg.

Klaus von Trotha, the state science

minister, said last week that the proposed legislation would allow the authorities to combat fraud and serious misconduct more rapidly and efficiently.

NIH should 'pay market rates to investigators'

[WASHINGTON] The US National Institutes of Health (NIH) should boost the \$125,000 ceiling on investigators' salaries "to appropriate market levels", says a report on the impact of managed care on academic health centres (AHCs).

The report by the Commonwealth Fund, a New York-based philanthropic think-tank that studies biomedical research issues, says that NIH is failing to pay enough in both direct and indirect costs to AHC-based investigators.

It calls on NIH to pay more for construction and renovation at AHCs, and says that the biomedical agency should fund grants at levels much closer to what applicants request. This, it says, would relieve AHCs of the burden of making up the difference.

Such cost sharing "may result in the allocation of research dollars on the basis of the recipient's ability to pay rather than strictly on the merits of the proposed work", warns the report.

Physicists lobby Jospin over institute vacancy

[PARIS] Nuclear physicists at France's Centre National de la Recherche Scientifique (CNRS) have written to Prime Minister Lionel Jospin asking him to insist that science minister Claude Allègre appoint a director for the agency's National Institute for Nuclear and Particle Physics (IN2P3). The institute has lacked a director since last October.

The letter contests the apparent desire of the science ministry to wait until the outcome of discussions over the future of IN2P3 before appointing a successor to former director Claude Detraz, who left to take up a post at the European Laboratory for Particle Physics (CERN) in Geneva, Switzerland.

One idea suggested is to transfer IN2P3 to the atomic energy commission. Another is to create a single physics department within the CNRS combining IN2P3 and the Department of Physical Sciences and Mathematics.

Japan-UK joint venture could seek rice varieties

[TOKYO] Japan Tobacco and the UK life sciences company Zeneca are discussing

setting up a joint venture by the end of the year to develop genetically modified agricultural products in Japan.

The aim is to develop new varieties of rice with the help of Zeneca's expertise in agricultural biotechnology. Japan Tobacco is the only Japanese private company participating in the Rice Genome Project, an international programme aiming to analyse the rice genome by 2010. The company is currently developing genetically modified rice for brewing sake and for use in livestock feed.

Blue skies ahead for climate collaboration

[LONDON] Climate researchers from Japan and the European Union (EU) ended a four-day workshop in Japan this month with an agreement to continue joint research, in particular in the area of high resolution and regional climate change simulations.

Researchers at the meeting also agreed to cooperate in collecting and analysing surface and sub-surface data from the Indian Ocean, both from satellites and from *in situ* observations. They recommended setting up an EU-Japan working group to maintain scientific collaboration between researchers from the two regions.

Supermarkets take GM foods off the shelf

[LONDON] Seven supermarket chains in Europe have joined forces to eliminate genetically modified ingredients from their own-label food products.

The initiative is being led by Sainsbury's, Britain's second largest chain, in response to last month's media controversy over genetically modified potatoes.

A Sainsbury's spokesman says the company has had 12,000 telephone calls

and numerous letters and e-mails in the past month from customers who are concerned about the issue.

In 1996, Sainsbury's introduced a genetically modified tomato purée, which outsold its non-GM counterpart until the controversy broke. The group's chairman at the time was Britain's science minister Lord David Sainsbury.



nature

the international weekly journal of science

invites applications for the post of

Chief Editor, Biological Sciences

Do you have what it takes to lead a growing team of talented editors publishing the world's finest biological sciences papers? To take on the most demanding and stimulating editorial role in biological sciences publishing? To be responsible for the highest standards in assessment and review? To represent *Nature* in the biological sciences community – and to ensure that *Nature* serves the interests of that community?

If so, we want to hear from you.

The Chief Editor, Biological Sciences, is a senior appointment, reporting directly to the Editor, backed by a strong commitment to investment and expansion. He or she will play a key part in developing *Nature's* publishing policy. Terms and conditions will be highly competitive, reflecting the critical importance and responsibilities of the role.



The ideal candidate will probably have some editorial experience, but will certainly have intellectual vision and strong leadership qualities. He or she will have significant achievements in biological sciences research; demonstrable interests across a wide range of biological sciences; respect in the scientific community; a capacity for prompt and balanced critical assessment; good communications skills; and imaginative ideas for the future development of *Nature* and their contribution to it.

The job is based in *Nature's* modern London offices but will involve international travel to meetings and laboratories.

Applicants should write in confidence, marked 'Personal', to:

Dr Philip Campbell,
The Editor, *Nature*,
4 Crinan St, London N1 9XW, UK.

They should include a CV, a discussion of relevant experience, and an indication of their specific ambitions for *Nature*. Applications should be sent as soon as possible and no later than 16 April.

Researchers face 'Catch-22' grants trap

Sir — In an era of shrinking funding of basic scientific research, increased competition has resulted in a sharpening of the peer-review process, with more elements coming into play during the 'sieving' of grant applications. In addition to the traditional criteria of originality, adequate methodology, relevance, publication record and so on, an increasingly important element is 'preliminary data'. The applicant is required to provide original, unpublished data that provide leads and justify the hypotheses of the proposal and give an indication of the expected results.

This element of the review process is amounting to alarming proportions. A colleague in the United States says he would not dare to submit a proposal to the National Institutes of Health (NIH) to study a certain developmental process in transgenic mice without having generated the animals and provided a 'preliminary' characterization of their phenotype, which in practice means having as much data on the creatures as possible, preferably of the kind that come close to answering the original question.

Who is supposed to pay for generating those transgenic mice? Presumably either 'leftovers' from previous grants or a generous collaborator. As it happens, being a renowned scientist, my colleague has access to both. But young assistant professors are trapped in a 'Catch-22' situation: they spend

so much time rewriting grant proposals that they never manage to gather the preliminary data requested by the reviewers. After having an NIH grant application returned for lack of *in vivo* work, another colleague submitted a revised application incorporating experiments in relevant animal models. The application came back again from the reviewers indicating that preliminary data from *in vitro* experiments would strengthen the case for the proposed *in vivo* work. For the third round, my colleague had to perform the experiments she had applied for in the first place!

How does one get that precious first piece of preliminary data? Like professional snooker players trying to keep the break going, scientists nowadays have to balance their experiments carefully between those that answer the questions originally raised and those that generate enough preliminary data to allow for another round of funding.

Fortunately, there remain some European funding agencies to which one can submit off-the-beaten-path or moderately risky ideas without having already performed all the crucial experiments. But, as competition for limited funds increases, the importance of preliminary data in the peer-review process is likely to rise among these sources as well. Soon, agencies all over the world will only be paying the few overdue bills remaining

from work that has already been done!

One dangerous consequence of this global trend is that only the most wealthy labs can afford to embark on truly innovative projects, study unexplored avenues or pursue seemingly odd leads — those that sometimes bring breakthroughs. Young scientists may have to content themselves with tightening the nuts and bolts of the established conceptual edifice. Is this what we really want for the future of basic research?

I propose the formation of an international funding agency, the Preliminary Data Organization — 'Predator'. This will be the first agency devoted to supporting the generation of preliminary data. It will consider only truly new ideas, and will disqualify grant applications that include preliminary data, on the principle that prior art nullifies originality. Predator will provide funds for the generation of all the necessary reagents, and for the execution of a minimal set of crucial experiments. Grant recipients will be able to use their data to apply to the more traditional organizations for funds to reproduce their findings in a closely related species! Any offers from rich philanthropists to take up this suggestion?

Carlos F. Ibáñez

*Division of Molecular Neurobiology,
Karolinska Institute, 17177 Stockholm, Sweden*

It's time to stop counting beans

Sir — The letter by Michael J. Larkin about research productivity among young UK scientists highlights a 'bean counter' mentality all too common in the academic community, apparently on both sides of the Atlantic (*Nature* 397, 467; 1999).

While compiling an analysis of the productivity and quality of research of a randomly selected group of US academic researchers funded by a large government agency ($n = 60$), we noted an almost complete lack of association between the number of years of experience in research and productivity as measured by the mean number of papers published per year ($R^2 = 0.103$; see Fig. 1).

This finding appears somewhat at odds with that of Larkin who showed that, within his cohort, mean numbers of papers increased steadily over the course of several years. This may reflect a difference in the publication patterns of UK versus US academic researchers at different career

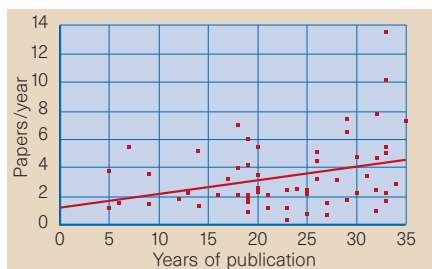


Figure 1 There is no strong relationship between experience in biomedical research (as measured by years of publication) and productivity (as judged by the number of peer-reviewed papers published per year).

stages. Even so, Larkin's evaluative criterion of a requirement for 'excellence' in the early career publications of the UK researchers in his study was not further defined.

Our assessment, using a quantifiable metric, clearly shows that the quality of research bears no relationship to the number of years of experience a researcher has accumulated in scientific publication ($R^2 = 0.0007$). Put another way, we believe our data suggest that excellence in scientific

research is manifested from the beginning of one's research career, and that publication rate (at least in the United States) does not change dramatically over the course of one's career. Therefore, perhaps more attention should be paid to where one's papers are published and less to how many papers are published.

David A. Watson

*National Space Biomedical Research Institute
and Department of Microbiology and Immunology,
University of Texas Medical Branch,
Nassau Bay, Texas 77058, USA*

Franklin recalled

Sir — I am writing a biography of the crystallographer Rosalind Franklin (1920–58) and would like to hear from anyone who knew or worked with her in London or Paris, or who met her, even briefly, at conferences or on holiday in Europe, Israel or the United States.

Brenda Maddox

*9 Pitt Street, London W8 4NX, UK
e-mail: bmaddox@pitt.demon.co.uk*

Is science dangerous?

Does society need protecting from scientific advances? Most emphatically not, so long as scientists themselves and their employers are committed to full disclosure of what they know.

Lewis Wolpert

The idea that knowledge is dangerous is deeply embedded in our culture. Adam and Eve were forbidden to eat from the biblical Tree of Knowledge, and in Milton's *Paradise Lost* the serpent addresses the Tree as the "Mother of Science". The archangel Raphael advises Adam to be "lowly wise" when he tries to question him about the nature of the Universe. Indeed, western literature is filled with images of scientists meddling with nature, with disastrous results. Scientists are portrayed as a soulless group, unconcerned with ethical issues.

But is science in fact dangerous, and do scientists have special social responsibilities? It is essential to recognize that reliable scientific knowledge has no moral or ethical value. Science tells us how the world is: that we are not at the centre of the Universe is neither good nor bad, nor is the possibility that genes can influence our intelligence or behaviour.

Moral obligations

Dangers and ethical issues come into play when scientific research is done in practice, for example in experiments involving humans and other animals, or when science is applied to technology, or in issues related to safety. There is thus an important distinction between science and technology: between knowledge and understanding on the one hand, and the application of that knowledge to making something, or using it in some practical way, on the other.

Science produces ideas about how the world works, whereas the ideas in technology result in usable objects. Technology is much older than science and, unaided by any science, it gave rise to early crafts such as agriculture and metalworking. I would argue that science made virtually no contribution to technology until the nineteenth century — even the great triumphs of engineering such as the steam engine and Renaissance cathedrals were built with imaginative trial and error, virtually without any impact of science¹.

Whatever new technology is introduced, it is not for scientists to make moral or ethical decisions about its use, as they have no special rights or skills in this regard. There is grave danger in asking scientists to be more socially responsible if they would also be given the right and authority to make such decisions on their own. The social obligations that scientists have, as distinct from those responsibilities they share with all citizens (such as supporting a democratic soci-

ety and taking care of the rights of others), come from them having access to specialized knowledge of how the world works that is not easily accessible to others. Their obligation is to make public any social implications of their work and its technological applications, and to give some assessment of its reliability. In most areas of science it matters little to the public whether a particular theory is right or wrong, but in some areas, such as human and plant genetics, it matters a great deal.

When the facts are examined dispassionately, it is not easy to find cases where scientists have behaved unethically in relation to the public. Contrary to some claims, there is no evidence that they did so either in the case of bovine spongiform encephalopathy (BSE) in the United Kingdom and elsewhere or in the AIDS blood scandal currently reverberating in France, for example.

The most clear case of immorality in scientific research was the eugenics movement². The scientific assumptions behind this were crucial: that most human attributes (desirable and undesirable) are inherited. The scientists concerned completely failed to give an assessment of the reliability of their ideas or sufficiently to consider their implications. On the contrary, and even more blameworthy, their conclusions seem to have

been driven by what they saw as the desirable social implications. In contrast, the Allied scientists who built the atomic bomb behaved morally, and fulfilled their social obligations by informing their governments about the implications of atomic theory. The decision to build the bomb was taken by politicians, not scientists³. Should scientists on their own ever be entitled to make such decisions? For the German eugenicists, there should have been a conflict between their responsibilities as scientists and as citizens.

How, then, should scientists behave when faced with a conflict between their responsibilities as researchers and their responsibility to those for whom they work? Should a scientist in government employment allow his or her superiors to keep the dangers of eating certain foods secret from the public? Similarly, what is the ethical position of a scientist working for a chemical company who believes a product is dangerous, yet whose employment contract requires confidentiality about the nature of the research? In both cases, one should not underestimate the problems in hazard assessment, in itself a complex business. The problem is no different to that of anyone, for example an accountant, who discovers corruption: if no action is taken after reporting the matter to his or her superiors, the individual must make a



DAVID NEWTON

very difficult decision. Scientists, just like everyone else, have to try not to become the unquestioning tools of their employers.

Genetic pornography

The very term 'genetic engineering' conjures up the image of Frankenstein and his monster — Mary Shelley was the unintentional evil fairy godmother of genetics — a tradition well-known in literature (*Brave New World*, *The Island of Dr Moreau* and so on), and most recently manifested by the likes of "Jurassic Park" and "Godzilla". The media are aware of this and often report what I regard as genetic pornography — reports dressed up to titillate and frighten. A nasty example was a widely disseminated picture of a mouse with a 'human' ear on its back — not a human ear at all but a piece of cartilage-like material. Newspapers print sensational and unjustified headlines such as the "Frankenstein foods" idiocy surrounding genetically modified organisms in the United Kingdom.

To apply genetic engineering requires considerable knowledge and, even more important, money, which in many cases is hard for scientists to come by. Indeed, for the public sector the expense of the applications of genetics and molecular biology can open up difficult choices: new medical treatments, requiring complex technology, cannot be given to all. There has to be some principle of rationing, and this poses serious moral and ethical dilemmas much more worthy of consideration than those of genetic engineering and the like.

Dangers of genetics

So what dangers does genetics pose to society? 'Bioethics' is a growth industry that purports to address this question, but one should regard this field with caution, as bioethicists have a vested interest in finding difficulties. Nevertheless, it has made some valuable contributions, including advice on experiments on human embryos in the United Kingdom and on the rights of fetuses. But advances in genetics raise few new ethical issues — there are no new ethical issues in relation to the current hysteria over cloning.

Some of the common fears about cloning are little more than science fiction at present, for example the danger of producing enormous numbers of genetically identical individuals. It is amusing to watch moralists swing from denying that genes have an important effect on intelligence or behaviour to saying that a cloned individual's behaviour will be entirely determined by the individual's genetic make-up. At present, the risk of human cloning leading to abnormalities is high and so it should not be attempted,

and I hope no mother would be so unwise as to become involved. Gene therapy — introducing genes to cure a genetic disease such as cystic fibrosis — has risks, as do all new medical treatments. There may well be problems with insurance and testing, but are these any different from those related to someone considered to be at increased risk of contracting AIDS or cancer?

Genetically modified foods have raised extensive public concern, and there seems no alternative but to rely on regulatory bodies to assess their safety (as is the case with other foods). The consumer is entitled to make a choice, and making a satisfactory choice requires trust or knowledge. But that depends on everyone sticking to the rules on quality control and full disclosure of what is in the food; the role of legislators is to make sure that these rules are rigorously followed. As with the licensing of medicines, each new genetically modified food must be considered individually. Science commissioned by a government and carried out in-house in government research labs is not appropriate when the results have important implications for public health and for government policy. It is essential in doing science to expose all one's acquired knowledge to criticism by others. The main lesson to be learned from the experience with BSE is that openness is all important.

Other fears relate to the so-called tyranny of knowledge which, claims Ian Kennedy, arises through the choices it forces on us "for which none of us is prepared spiritually or intellectually"⁴. Thus, couples may be faced with difficult choices about prenatal diagnosis of genetic diseases: this can lead to choices about whether or not to terminate a pregnancy, or whether to inform siblings of a possible genetic risk of which they are not aware. There are problems, but I believe that one must not underestimate people's capacity to deal with difficult choices when they understand the issues. Ultimately, the choice as to whether to seek knowledge rests with the individual. Most ethical issues in medicine are best resolved by considering the rights of the people involved to determine their own future.

Censorship

Are there areas of research that are so socially sensitive that they should be avoided, even proscribed? One possible area is the genetic basis of intelligence, and particularly the possible link between race and intelligence. Are there, as the literary critic George Steiner has argued, "certain orders of truth which would infect the marrow of politics and would poison beyond all cure the already tense relations between social classes and these communities"? In short, are there doors in front of current research that should be marked "Too dangerous to open"?

I realize the dangers, but I cherish the openness of scientific investigation too much

to put up such a notice. I stand by the distinction between knowledge of the world and how it is used. So I must answer Steiner's question in the negative, provided of course that scientists fulfil their social obligations. The better understanding we have of the world, the better chance we have of making a just society. One should not abandon the possibility of using a scientific idea to do good because one could use the same idea to do bad. There is no knowledge that is not susceptible to manipulation for evil purposes.

Once one begins to censor the acquisition of objective knowledge, one is on the most slippery slope of all. Scientists cannot easily predict the social and technological implications of research. It was once argued that radio waves would have no practical applications, and Lord Rutherford famously said that the application of atomic energy was moonshine. Those investigating the resistance of certain bacteria to viral infection did not predict the discovery of restriction enzymes, an indispensable tool for cutting up DNA and hence the basis of genetic engineering.

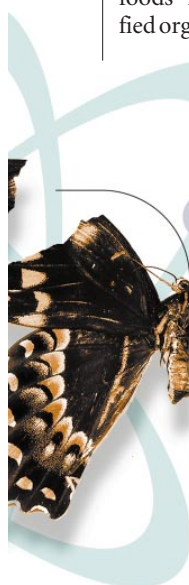
To those who doubt whether the public or politicians are capable of making the 'correct' decisions about science and its applications, I commend the advice of Thomas Jefferson: "I know no safe depository of the ultimate powers of the society but the people themselves, and if we think them not enlightened enough to exercise that control with a wholesome discretion, the remedy is not to take it from them, but to inform their direction."

But how do we ensure that the public are involved in decision-making, and that scientists, doctors, engineers, bioethicists and other experts, who must be involved, do not appropriate decision-making for themselves? How do we ensure that scientists take on the social obligation of making the implications of their work public? We must rely on our democratic institutions: elected representatives; a free, vigorous and even responsible media; affected groups; and the researchers themselves. National and international councils that can assess the ethical issues relating to the applications of science and promote public debate are no doubt valuable. But one wonders what such a committee would have said if the public had been offered a convenient form of transport, but at the cost, in the United Kingdom alone, of more than 3,000 lives per year, a quarter of a million injured and the untold damage of pollution. Where are the car-ethicists? □

Lewis Wolpert is in the Department of Anatomy and Developmental Biology, University College London, Gower St, London WC1E 6BT, UK.

e-mail: l.wolpert@ucl.ac.uk

1. Wolpert, L. *The Unnatural Nature of Science* (Faber, London, 1992).
2. Kevles, D. J. *In the Name of Eugenics* (Univ. California Press, Berkeley, 1985).
3. Rhodes, R. *The Making of the Atomic Bomb* (Simon & Schuster, New York, 1986).
4. Kennedy, I. in *Predicting the Future* (eds Howe, L. & Wain, A.) 96–117 (Cambridge Univ. Press, 1993).



At the roots of the mammalian family tree

Timothy Rowe

An exceptionally complete skeleton, dating back roughly 140–150 million years, offers our closest look yet at the last common ancestor of modern mammals.

Mesozoic mammals are among the most challenging and prized entries on a fossil hunter's list of discoveries. The essence of the challenge is size — the Mesozoic history of mammals was played out by tiny animals. Few specimens survived the destructive agencies of fossilization, and those that did are supremely difficult to find and collect. Most are fragmentary, and most named species are based on isolated teeth and jaws. In nearly two centuries of searching, only a few precious complete specimens have been recovered and, without better fossils, long stretches of mammalian history have remained in the dark. But, on page 326 of this issue, Ji, Luo and Ji¹ describe one of the most complete and exquisitely preserved specimens ever found. It comes from the same Late Jurassic/Early Cretaceous deposit of Liaoning, China, that recently yielded spectacular feathered dinosaurs² and one other complete mammal skeleton³. The latest discovery helps to fill a wide gap in the fossil record, and brings new information to classic problems on the origin and interrelationships of early mammals.

The first Mesozoic mammals were discovered in 1812 by a mason in a tilestone

quarry near Headington, England. These specimens came from the Middle Jurassic (roughly 165-Myr-old) Stonesfield Slate, and consisted of two isolated lower jaws, each belonging to a different species. Today, the Stonesfield jaws are still the oldest known fossils of the 'crown clade' Mammalia⁴ — the lineage founded by the last common ancestor of living mammals (Fig. 1).

The tiny jaws quickly found their way to the University of Oxford where, in 1818, Baron Georges Cuvier examined them while on a sojourn in England. Renowned for his ability to judge the nature and affinities of an extinct animal from a part or even a single fragment of a skeleton, Cuvier pronounced the Stonesfield specimens to be mammalian⁵. He lived up to his reputation, and his identification was the first of many violations of what had been considered a very general rule — that mammals did not live during the Age of Reptiles.

A century and a half later and hundreds more Mesozoic mammal fossils had been discovered, yet the Stonesfield jaws remained among the most complete specimens known. Screen-washing techniques pioneered by Claude Hibbard in the 1940s offered the first clues to an unsuspected diversity of Mesozoic mammalian species⁶.

Many tonnes of Mesozoic sediments were sieved through a series of screens designed to trap even the smallest fossils. But most recovered specimens consisted only of teeth and broken jaws, and the emerging view of early mammalian history became overly focused. In 1968, Alfred Sherwood Romer⁷ admonished: "So great has been this concentration on dentitions that I often accuse my 'mammalian' colleagues, not without some degree of justice, of conceiving of mammals as consisting solely of molar teeth and of considering that mammalian evolution consisted of parent molar teeth giving birth to filial molar teeth and so on down through the ages."

The first great advance towards a more complete knowledge of the structure and relationships of Mesozoic mammals came in the 1960s, when the Polish Academy of Sciences sent a series of expeditions into central Asia. Dozens of Late Cretaceous mammal skeletons representing several different lineages were collected⁸. Throughout the 1990s, Asian expeditions led by the American Museum have been collecting hundreds more Late Cretaceous specimens that document, in even greater detail, the initial diversification of therian mammals (Fig. 1)^{8,9}.

Computer-assisted cladistic analyses of data from these more complete specimens^{4,8,9} profoundly altered the picture of Late Cretaceous mammalian diversity that was painted in Romer's time¹⁰. For example, Romer's generation accepted that Mammalia arose in the Triassic (which immediately preceded the Jurassic), whereas the new analyses indicate that the last common ancestor of living mammals probably lived in the Early or Middle Jurassic. In other words, Mammalia is 20–40 Myr younger than once believed. Until very recently, however, the earliest details of mammalian history were obscured owing to the lack of complete fossils.

The remarkable specimen described by Ji and colleagues¹ is, along with a primitive therian mammal announced last year from the same locality³, by far the most complete and informative fossil discovered from a roughly 20-Myr or longer segment of Jurassic and Early Cretaceous time. Ji *et al.* present an analysis of evolutionary relationships, including dental evidence and data from throughout the skeleton, that places the new find very near the base of the mammalian crown clade (Fig. 1). Their analysis indicates that Triconodontidae — a group once believed to contain the direct ancestors of modern mammals — is not a natural group. Originally founded on the dental attributes that inspired its name, some triconodonts seem to be closer to mammals than to other so-called triconodonts.

But Ji and colleagues' specimen also highlights the presence of homoplasy — the independent evolution of similar features

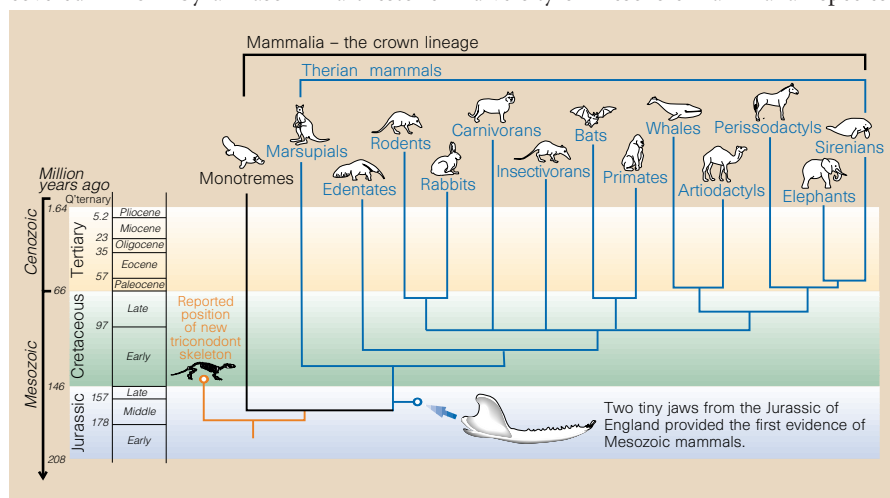


Figure 1 Timescale of mammalian evolution. The new skeleton discovered by Ji *et al.*¹ has been placed near the base of the crown lineage, Mammalia.

— in skeletal characters, corroborating an earlier finding that no region of the skeleton is immune to homoplasy¹¹. Still, the data to be gleaned from the skeleton are strong enough to overthrow the apparent dental resemblance of the triconodonts to one another. And it is the skeletal characteristics that largely support the sister-group relationship of multituberculates (a long-lived and long-enigmatic lineage of extinct mammals) with therian mammals, once again contradicting hypotheses derived from dental evidence.

This beautiful specimen also offers new insight into what the ancestor of modern mammals was like. Working out when mammals first moved into the trees, and whether this happened more than once, has been problematic. Ji and colleagues' find indicates that mammals arose as terrestrial forms, and that only later did their therian descendants take to the trees.

Even with this spectacular new find, long gaps still punctuate our Mesozoic record of mammals and their extinct relatives. But this exciting Chinese locality has now produced

so many exquisite tetrapod fossils that additional complete specimens of early mammals are likely to be unearthed. We can then expect rapid increases in the resolution of what was once the most fragmented segment of our early history. □

Timothy Rowe is in the Department of Geological Sciences and the Vertebrate Paleontology Laboratory, University of Texas at Austin, Austin, Texas 78712, USA.
e-mail: rowe@mail.utexas.edu

1. Ji, Q., Luo, Z. & Ji, S.-A. *Nature* **398**, 326–330 (1999).
2. Ji, Q., Currie, P. J., Norell, M. A. & Ji, S.-A. *Nature* **393**, 753–761 (1998).
3. Hu, Y., Wang, Y., Luo, Z. & Li, C. *Nature* **390**, 137–142 (1997).
4. Rowe, T. *J. Vert. Paleontol.* **8**, 241–264 (1988).
5. Owen, R. *Geol. Trans* (2nd series) **6**, 47–65 (1839).
6. Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A. *Mesozoic Mammals — The First Two-thirds of Mammalian History* (Univ. California Press, Berkeley, 1979).
7. Romer, A. S. *Notes and Comments on Vertebrate Paleontology* (Univ. Chicago Press, 1968).
8. Novacek, M. J. *et al. Nature* **389**, 483–486 (1997).
9. Rougier, G. W., Wible, J. R. & Novacek, M. J. *Nature* **396**, 459–463 (1998).
10. Dingus, L. & Rowe, T. *The Mistaken Extinction — Dinosaur Evolution and the Origin of Birds* (Freeman, New York, 1997).
11. Rowe, T. in *Mammalian Phylogeny* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 129–145 (Springer, New York, 1993).

proteins⁵. The picosecond X-ray barrier was eventually broken^{6,7} following rapid developments in high-power laser technology (see Box 1).

On page 310 of this issue Rose-Petrucci *et al.*⁸ describe how they generate picosecond bursts of copper K_{α} radiation by irradiating a thin copper wire with a short-pulse high-power laser. They then use this ultrashort radiation to measure the response of a gallium arsenide (GaAs) crystal to sudden heating. This is a remarkable achievement, not only in terms of the science, but also because this picosecond X-ray diffraction did not require a large facility, but used equipment of a cost and scale commensurate with the ambitions of a well-equipped university department. This work highlights a burgeoning field of science, which may ultimately allow changes in electron density to be monitored during biological and chemical reactions, with femtosecond resolution.

This particular paper brings together X-ray diffraction and picosecond ultrasonics. When a femtosecond laser pulse of suitable intensity is incident on an absorbing material, a thin layer at the surface is heated, but the heating is so rapid that it takes place before the layer is able to expand. So, the hot region, still at its initial density, is at a high pressure — typically a few kilobars if the layer is near melting point. The material then relaxes as an acoustic wave travels into the material, forming regions of both expansion and compression (see Fig. 3 on page 312). Usually the pressure pulses are monitored by reflecting optical light from the surface, enabling the detection of underlying structures such as defects in computer chips⁹, just as conventional sonar can detect underwater objects. Instead, Rose-Petrucci *et al.*⁸ use picosecond pulses of X-rays to directly monitor the spacing between the atoms in the heated layer as a function of

X-ray diffraction

Table-top picosecond sources

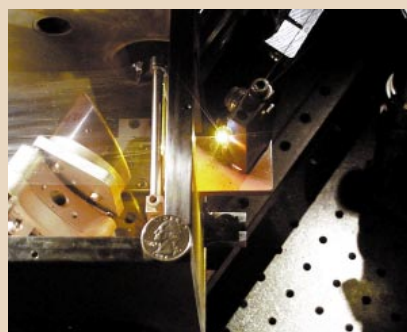
Justin Wark

X-ray diffraction has long been one of the most widely used diagnostic tools in both the physical and life sciences. It can be used to determine structures of materials as simple as a grain of salt, and as complex as a virus or protein in the crystalline form. The majority of studies, using either conventional electron-impact sources or synchrotron radiation, yield information about the static structure of a sample. However, there has long been a desire for ultra-short X-ray pulses in order to follow the

evolving electron density of structures as they are altered. Somewhat surprisingly, X-ray diffraction patterns with millisecond exposures were first produced over half a century ago¹, and for the past decade bunches of electrons circulating within synchrotrons², and plasmas created with large laser systems, have been used to generate X-ray pulses with tens of picoseconds to nanosecond duration. Such pulses have already been used to study shocked and annealed crystals^{3,4}, and photo-initiated reactions within

Box 1: Brighter, faster, smaller

Advances in high-power laser technology led to the development of titanium:sapphire lasers that can generate approximately 1 J of 800-nm-wavelength light in a pulse length of about 10 femtoseconds at a repetition rate of 10 Hz (refs 11–14). These peak laser powers ($> 10^{13}$ W) are truly enormous — greater, even, than the electrical power output of the whole planet at any instant. When the output from such a laser is focused onto a target, a plasma is formed, and the laser light is absorbed in this plasma up to a point where the laser frequency equals the natural frequency of oscillation of the plasma. At this resonant position, plasma waves are



driven to such large amplitudes that they break, releasing the electrons within them at high energies. These electrons penetrate the underlying solid material and generate K_{α} radiation, just as they would do in a standard X-ray

tube. Present studies indicate that the X-ray pulse can be less than a picosecond in duration, but is usually longer than the laser pulse itself owing to both the time taken for the electrons to penetrate the solid, and their complex trajectories in the magnetic fields produced by such large electron currents. The figure shows the vacuum chamber in which the X-rays used by Rose-Petrucci *et al.*⁸ were produced — a bright flash of optical light can be seen from the plasma emitted by the laser-irradiated copper wire. The X-ray source itself is less than a hair's breadth in diameter, and is confined to a thin layer at the surface of the wire.

J. W.

both depth and time. So they directly observe the evolution of both the expansion and compression waves with picosecond resolution.

The authors use a titanium:sapphire laser both to heat the test sample, and synchronously to produce the X-ray pulse (see Box 1). By use of a beamsplitter, a small part of the laser energy irradiates a GaAs wafer to produce the ultrasonic pressure pulse, while the rest is focused on a small spot of the copper wire to produce the X-rays. These picosecond X-rays are then diffracted from the GaAs crystal and recorded on a charge-coupled device. The exact timing of the X-ray diffraction from the GaAs can be altered by changing the optical path difference between the two parts of the beams. This allowed Rose-Petruck *et al.* to monitor the acoustic pulse for several hundred picoseconds after it was initiated. The acoustic waves caused by the ultrasonic pressure within the crystal are monitored by recording the change in the Bragg scattering angle owing to the change in the distances between the atoms. As the X-rays simultaneously penetrate regions of both compression and expansion, a complicated algorithm is used to extract the strain as a function of depth.

Further developments in time-resolved diffraction are on the horizon. Synchrotron radiation with a timescale of a few pico-

seconds has been used to monitor laser-irradiated semiconductors¹⁰, and sub-picosecond X-ray pulses have been generated by Thomson scattering lasers from high-energy electron beams⁷. Such sources, only available at large facilities, complement the table-top K_{α} sources, as they can produce broadband X-radiation. But, whatever the relative merits of synchrotron versus laser-generated sources finally turn out to be, the work of Rose-Petruck *et al.*⁸ clearly demonstrates the impressive potential of table-top picosecond X-ray sources, and brings us closer to the goal of watching, on femtosecond timescales, so-called 'molecular movies'. □

Justin Wark is in the Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

e-mail: justin.wark@physics.ox.ac.uk

1. Tsukermann, V. A. & Andeenko, A. I. *Zh. Tekh. Fiz.* **12**, 185–194 (1942).
2. Szebenyi, D. M. E. *et al.* *J. Appl. Cryst.* **25**, 414–423 (1992).
3. Larson, B. C., White, C. W., Noggle, T. S. & Barhorst, J. F. *Appl. Phys. Lett.* **42**, 282–284 (1983).
4. Wark, J. S. *et al.* *Phys. Rev. B* **35**, 9391–9394 (1987).
5. Srajer, V. *et al.* *Science* **274**, 1726–1729 (1996).
6. Rischel, C. *et al.* *Nature* **390**, 490–492 (1997).
7. Schoenlein, R. W. *et al.* *Science* **274**, 236–238 (1996).
8. Rose-Petruck, C. *et al.* *Nature* **398**, 310–312 (1999).
9. Maris, H. *Sci. Am.* **278**, 86–89 (1998).
10. Larsson, J. *et al.* *Appl. Phys. A* **66**, 587–591 (1998).
11. Bart, C. P. J. *et al.* *Opt. Lett.* **21**, 668–670 (1996).
12. Chambaret, C. *et al.* *Opt. Lett.* **21**, 1921–1923 (1996).
13. Yamakawa, K. *et al.* *Opt. Lett.* **23**, 1468–1470 (1998).
14. Backus, S. *et al.* *Rev. Sci. Instrum.* **69**, 1207–1223 (1998).

Olfaction

Good reception in fruitfly antennae

Yitzhak Pilpel and Doron Lancet

Insects have exquisite chemosensory faculties^{1–3} — some can sense pheromones from several miles away. Yet although the genes for olfactory receptors in vertebrates^{4,5} and the nematode worm *Caenorhabditis elegans*⁶ have been known for years, their insect counterparts have remained elusive. Now, reports in *Neuron*⁷ and *Cell*⁸ describe the identification of genes for olfactory receptors in the fruitfly *Drosophila melanogaster*. These genes encode membrane receptor proteins that probably mediate odorant recognition in the fly.

The discovery of *Drosophila* olfactory receptor genes was possible largely due to the availability, in databases, of genomic DNA sequences. To find these genes, Clyne *et al.*⁷ initially used a pure *in silico* approach. They assumed that *Drosophila* olfactory receptors would be structurally similar to the other known olfactory receptor genes. So, the authors identified two candidate genes from databases with an algorithm based on seeking their potential to code for proteins with multiple transmembrane segments. Vosshall *et al.*⁸ first identified an olfactory-specific rare messenger RNA (as befits a protein expected to be expressed in only a small sub-

set of sensory neurons). Both groups then found further transmembrane homologues by searching the archives of the *Drosophila* genome project and they confirmed that, as expected, the dozen or so candidate genes are specifically expressed in sub-populations of cells within the fly olfactory organs — the antennae and maxillary palps.

On statistical grounds, the new studies indicate that between 100 and 200 genes may code for the *Drosophila* olfactory receptors. This is far fewer than the estimated 500–1,000 genes active in many vertebrates, and even in the lowly *C. elegans*. On the other hand, the fruitfly may have a few dozen chemosensory cell types (compared with just 14 in the nematode), which converge on almost 50 vertebrate-like synaptic targets known as glomeruli².

Drosophila also seems to be the most ancient creature in which olfactory receptor genes are clonally excluded — that is, just one or a few genes are expressed in each of the sensory cells (Fig. 1, overleaf)^{7,8}. This is an important feature if each odorant is to be accurately represented across the sensory neuronal 'activation vector' (that is, the firing frequency input from different types of



100 YEARS AGO

It is interesting to note how the gradual discovery of the attendants of the various planets has influenced the compounding of the "laws" which from time to time have been found to approximately represent the positions of these bodies in the solar system. From the first discovery of Jupiter's four satellites by Galileo in 1610 to Huyghens, Cassini, and Sir W. Herschel, no regular relationship was perceived. When, however, in August 1877, Prof. Asaph Hall discovered the two moons of Mars, Deimos and Phobos ... it was seen that all the then known satellites were grouped in a geometrical progression, reckoning outwards from the Earth. Thus the Earth had one, Mars two, Jupiter four and Saturn eight. This seeming regularity was broken by the discovery on September 9, 1892, of a fifth satellite to Jupiter This last discovery of a ninth satellite for Saturn will furnish a reason for a new series being formed, as counting from the Earth outward from the Sun, the numbers of satellites to the planets Earth, Mars, Jupiter and Saturn are now 1, 2, 5 and 9 respectively, and these numbers are very nearly proportional to the distances of those planets from the Sun.

From *Nature* 23 March 1899.

50 YEARS AGO

Much of the published data regarding the mode of action of penicillin is necessarily concerned with secondary effects on bacteria, for example, rate of killing under different conditions, morphological changes, etc. By use of radioactive penicillin, it should be possible to find out directly something of the nature of the primary reaction between penicillin and the cell. Penicillin is taken up by all bacteria in amounts which increase with penicillin concentration There is, in fact, a direct correlation between the sensitivity of an organism and the amount of penicillin attached to it. ... However, if growth is halted by cooling, or if the cells are already dead, there is still a rapid but smaller uptake. ... Whatever the nature of the primary site of penicillin action, it is clear from this and other work that rapid growth exposes more centres in the bacteria with which the penicillin can react. The combination must either be by unusually strong adsorptive or by chemical forces.

From *Nature* 26 March 1949.

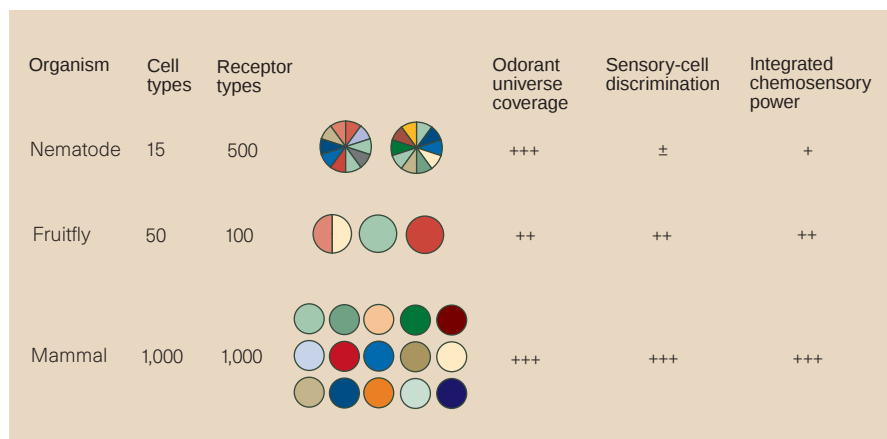


Figure 1 Evolution of olfactory cells and receptor proteins. Nematodes and mammals have similar numbers of chemosensory receptor types^{4,6}, so they should detect the same (large) number of odorants with comparable sensitivities¹³. Because, in the nematode, sensory cells each bear many different types of receptor, individual cells may show poorer discrimination among odours. But mammals seem to have complete clonal exclusion — each sensory cell expresses only one type of receptor, giving the best possible discrimination at the level of single neurons. The fruitfly may represent an intermediate evolutionary step; the number of cell types has increased modestly, but the sensory neurons have gone most of the way towards clonal exclusion. This means that, although its olfactory receptor repertoire is smaller, it is much better than the nematode at discriminating odours. Additional gene-repertoire enhancement is mainly what is needed to attain the olfactory power of mammals.

cell, used for integration of signals in the central nervous system)^{2,4}. The new studies even shed light on the mechanisms by which such cellular exclusion is controlled. Vosshall *et al.*⁸ suggest that a region 3 kilobases upstream of an olfactory receptor gene is involved, and Clyne *et al.*⁷ implicate a transcription factor⁹ called Acj6. In this respect, perhaps studies of the potentially simpler insect mechanisms will help us to work out how expression of olfactory receptor genes is controlled in higher organisms.

Could *Drosophila* and other insects have been the first to evolve the neuronal-integration device that vertebrates subsequently came to have²? If so, perhaps insects temporarily sacrificed the number of olfactory receptor genes so that they could evolve the elaborate cell-biological mechanisms needed to process chemosensory information more accurately (Fig. 1). Only later, with the advent of the vertebrate olfactory system, could both high receptor count and sophisticated neuronal wiring exist together. Interestingly, a dendrogram analysis (Fig. 2), which shows the relationship between different genes and organisms, may place the new *Drosophila* olfactory receptor genes, as well as some nematode chemosensory receptors, slightly nearer to receptor genes^{10,11} expressed in a vertebrate 'accessory olfactory pathway' (the vomeronasal organ, and the recently discovered¹² vertebrate taste receptors). So, the main olfactory system in vertebrates has emerged with a separate set of chemoreceptive genes, more akin to visual photoreceptors and neurotransmitter receptors, which may have taken over an existing clonal-exclusion machinery.

The *Drosophila* olfactory receptor genes are unique, and there are no obvious counterparts with similar sequences in other species. In fact, their sequence is so different from other chemosensory receptors that one could even suspect a transduction mechanism¹ other than the usual coupling to guanine-nucleotide-binding (G) proteins^{4,5}. In this, olfaction is unlike many other functional pathways, where clear structural and functional relationships can be traced across 500 million years of evolution (Fig. 2).

One possible reason for this is that fruitflies might have an idiosyncratic odour world⁸. To find out whether this is so, we could compare the olfactory-receptor sequences in insects with different behaviours and habitats. An alternative explanation — which is more consistent with studies that indicate nothing special about insect odorants³ — is that chemosensory repertoires evolved with a free rein, perhaps under balancing selection (which is known to generate diversity for diversity's sake). Just as in a combinatorial library, an olfactory repertoire could be fully functional as long as it is large and eclectic enough¹³. This could lead to the appearance of disparate chemosensory proteins in different phyla.

Some characteristics of invertebrate olfactory receptor genes contrast with those of the vertebrate receptors. For example, whereas the *Drosophila* and *C. elegans* genes are interrupted by introns within the protein-coding sequences, the coding regions of vertebrate olfactory receptor genes are intronless. This curious phenomenon suggests that the loss of introns is, in this case, a later evolutionary adaptation. Because many

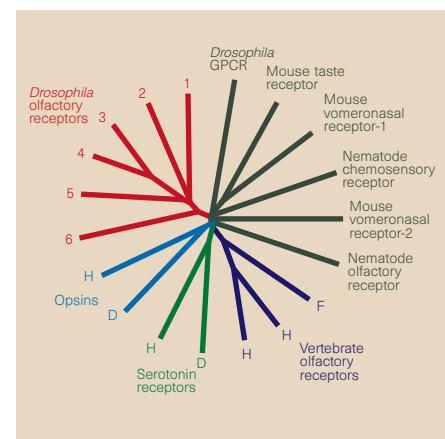


Figure 2 Dendrogram analysis for the protein sequences of chemosensory receptors and G-protein-coupled receptors. Included are six of the *Drosophila* olfactory receptors discovered by Clyne *et al.*⁷ and Vosshall *et al.*⁸. Although opsins and serotonin receptors are clear cases of human–*Drosophila* orthology, the *Drosophila* olfactory receptors are almost unrelated to their vertebrate counterparts. H, human; F, fish; D, *Drosophila*; GPCR, the G-protein-coupled (Frizzled-like) receptor.

other seven-transmembrane-domain receptors are also intronless, the phenomenon could reflect a general characteristic of this type of transduction protein¹⁴. However, the mechanisms leading to the loss of introns in olfactory-receptor genes may have had more to do with the need for extensive gene duplication^{5,14} to allow the evolution of fully fledged vertebrate olfactory repertoires.

In less than a year the *Drosophila* genome will probably be fully sequenced, leading to a complete knowledge of the fruitfly 'olfactory sub-genome'. By unravelling the olfactory-receptor genes in one of the most coveted model organisms, replete with powerful genetic and developmental tools, we will surely be able to solve many of the questions about how we, and other species, perceive the chemical universe that surrounds us. □

Yitzhak Pilpel and Doron Lancet are in the Department of Molecular Genetics and the Genome Center, Weizmann Institute of Science, 76100 Rehovot, Israel.

e-mails: bnpilpel@membran1.weizmann.ac.il
bmlancet@weizmann.ac.il

- Kaissling, K. E. *Chem-Senses* **21**, 257–268 (1996).
- Hildebrand, J. G. & Shepherd, G. M. *Annu. Rev. Neurosci.* **20**, 595–631 (1997).
- Carlson, J. R. *Trends Genet.* **12**, 175–180 (1996).
- Buck, L. B. *Annu. Rev. Neurosci.* **19**, 517–544 (1996).
- Lancet, D. & Ben-Arie, N. *Curr. Biol.* **3**, 668–674 (1993).
- Bargmann, C. I. & Kaplan, J. M. *Annu. Rev. Neurosci.* **21**, 279–308 (1998).
- Clyne, P. J. *et al. Neuron* **22**, 327–338 (1999).
- Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. *Cell* **96**, 725–736 (1999).
- Clyne, P. J. *et al. Neuron* **22**, 339–347 (1999).
- Dulac, C. & Axel, R. *Cell* **83**, 195–206 (1995).
- Matsunami, H. & Buck, L. B. *Cell* **90**, 775–784 (1997).
- Hoon, M. A. *et al. Cell* **96**, 541–551 (1999).
- Lancet, D., Sadovskiy, E. & Seidemann, E. *Proc. Natl Acad. Sci. USA* **90**, 3715–3719 (1993).
- Gentles, A. J. & Karlin, S. *Trends Genet.* **15**, 47–48 (1999).

Cosmology

The pulling power of galaxy clusters

Peter Coles

According to the textbooks, the large-scale Universe should be smooth and well behaved. Distant galaxies ought to be evenly distributed in space, and their motions should correspond to a pure 'Hubble flow', a uniform expansion of space in all directions. These are consequences of the 'cosmological principle', the theoretical notion that the Universe, in some average sense, is homogeneous (the same in every place) and isotropic (looks the same in every direction). But a study of the motions of distant galaxy clusters by Hudson *et al.*¹, appearing in the *Astrophysical Journal*, has found evidence of huge streaming motions that raises serious questions about how smooth the large-scale Universe can really be and puts theoretical models under severe pressure.

Although the cosmological principle is one of the central tenets of cosmological theory, it is obvious that the Universe is not exactly homogeneous and isotropic. Matter is not smoothly distributed, but organized into galaxies, galaxy clusters and even superclusters of galaxy clusters. This complex hierarchy of density fluctuations is, according to 'inflation theory', a result of the gravitational amplification of low-amplitude 'ripples' that were present in the very early Universe. But there should be a scale beyond which gravity has not had sufficient time to produce structures, and beyond which the Universe should therefore appear homogeneous. In this sense, the Universe should resemble a decorative carpet: on small scales one sees the details of the pattern, but overall it looks uniform as the elements repeat themselves.

Besides generating spatial patterns, gravity also generates velocities. In a perfectly uniform universe, everything moves away from everything else with a velocity v that is proportional to the distance between, d , such that $v = H_0 d$. This is known as the Hubble law, and the constant of proportionality H_0 is the Hubble constant. But the presence of density fluctuations distorts this uniform Hubble flow by introducing peculiar motions. For example, the Andromeda nebula (M31) is not receding from the Milky Way in line with Hubble's law. In fact, it is not receding at all, it is approaching. The gravitational effect of the Milky Way is strong enough to defeat the cosmological expansion, at least locally.

All galaxies execute some kind of peculiar motion, as a consequence of the gravitational influence of the lumpy distribution of

material around them. In the densest galaxy clusters — known as Abell clusters — where gravitational forces are very strong, galaxies rattle around like bees in a bottle, with peculiar velocities of $\sim 1,000 \text{ km s}^{-1}$ generated by the deep potential well in which they reside. These random 'thermal' motions are not the only forms of peculiar activity. On scales larger than individual clusters, the concerted action of entire superclusters produces a calmer, more coherent flow towards regions of above-average density, and away from regions of below-average density. These 'streaming' motions contain clues to the size of the largest structures doing the pulling and thus furnish an important test of cosmological models.

Measuring streaming motions is difficult, because one has to estimate the distance to an object before one can determine its peculiar velocity (the difference between its total velocity, measured using the Doppler shift of the spectrum, and the part due to the Hubble expansion, which depends on its distance). Systematic observational studies of streaming motions effectively began in 1976 (ref. 2). This was soon followed by the detection of a dipole anisotropy (a sinusoidal variation across the sky) in the cosmic microwave background, now attributed to the motion of our Galaxy and the Local Group of galaxies through the uniform radiation bath left behind by the Big Bang³.

In 1988, a study of streaming motions in a sample of elliptical galaxies revealed evidence for a systematic flow, simple modelling of which suggested that it could be explained by a hypothetical object about 60 megaparsecs (Mpc) away from the Milky Way⁴, which became known as the 'Great Attractor'. Although the Great Attractor model could indeed explain the observed flow pattern, it is neither the only nor even the most likely explanation for what was seen. The net effect of many larger but more distant structures seems a more reasonable picture. Whether this is the case or not, few doubted that deeper samples would eventually yield smaller streaming motions. After all, the Universe at large should be at rest with respect to the relic radiation field, shouldn't it?

In 1994, Lauer and Postman⁵ stunned the astronomical world by announcing the discovery of a huge bulk flow of a sample of galaxies in clusters. The sample depth was about 100 Mpc and the measured bulk flow almost 700 km s^{-1} . The direction of this anomalous flow is surprising too. The veloci-

ty of the Local Group with respect to the sample is not in the same direction as its velocity with respect to the relic microwave radiation. If the microwave background dipole is indeed due to motion through a 'cosmic rest frame', some enormous structure, beyond the depth of the Lauer-Postman sample, must be pulling the Local Group. Such a huge flow on such large scales and in such an unexpected direction is at odds with the idea that the Universe is smooth on large scales. There should be no concentrations of matter large enough to pull galaxy clusters about in this way. Is the Universe flouting the cosmological principle?

Hudson and colleagues¹, working on the Streaming Motions of Abell Clusters project (SMAC), have addressed this question. They use empirical correlations between the surface brightness, size and velocity dispersion of elliptical galaxies in clusters (the so-called fundamental plane) to construct a distance indicator, which they then apply to about 700 galaxies in 56 rich clusters. Similar to the Lauer and Postman sample, these objects lie in a region about 100 Mpc in radius and the sample is also participating in a massive bulk flow. A region containing about 10^{17} solar masses' worth of material, centred on the Local Group, is moving *en masse* at about 600 km s^{-1} through the Universe. The direction of the motion is, however, very different from that found by Lauer and Postman.

It is premature to regard the SMAC project as definitive, especially given the controversial history of streaming-motion studies in general, but it may represent the glimmer of an emerging consensus. From a theoretical point of view, the revised direction of flow makes much more sense when compared to the cosmic microwave background. Indeed, the direction is very close to that predicted on account of the observed distribution of distant X-ray clusters and the theory of how such motions are induced by matter fluctuations⁶. It also agrees with some related studies, in particular one by Willick⁷ based on spiral, rather than elliptical, galaxies.

Despite its large size, the direction of the SMAC flow considerably shortens the odds a theoretician like myself would lay on the flow being genuine. On the other hand, Dale *et al.*⁸, in another study of spiral galaxies, find a much smaller flow that contradicts both Willick and SMAC. If the SMAC flow is real, however, there must be concentrations of mass sufficiently large to generate the required pull. Most popular cosmological models do not produce such a lumpy Universe. Whether models can be made that do this, and at the same time satisfy all the other constraints arising from new observational methods, remains to be seen. □

Peter Coles is in the School of Physics & Astronomy, University of Nottingham, University Park,

Nottingham NG7 2RD, UK.

e-mail: Peter.Coles@Nottingham.ac.uk

1. Hudson, M. J., Smith, R. J., Lucey, J. R., Schlegel, D. J. & Davies, R. L. *Astrophys. J.* **512**, L79–L82 (1999).
2. Rubin, V. C., Ford, W. K. Jr, Thonnard, N., Roberts, M. S. & Graham, J. A. *Astrophys. J.* **81**, 687–737 (1976).

3. Smoot, G. F. & Lubin, P. M. *Astrophys. J. Lett.* **234**, L83–L86 (1979).
4. Lynden-Bell, D. *et al. Astrophys. J. Lett.* **326**, 19–49 (1988).
5. Lauer, T. R. & Postman, M. *Astrophys. J.* **425**, 418–438 (1994).
6. Plionis, M. & Kolokotronis, V. *Astrophys. J.* **500**, 1–7 (1998).
7. Willick, J. A. <http://xxx.lanl.gov/abs/astro-ph/9812470>
8. Dale, D. A. *et al. Astrophys. J.* **510**, L11–L14 (1999).

Climatology

Europe's winter prospects

Yochanan Kushnir

For northern Europeans, the winter of 1995–96 was a harsh reminder of the 1960s and early 1970s, when some feared the return of the Little Ice Age. Between November 1995 and February 1996 average temperatures were 2–4 °C below normal over a region stretching from Scandinavia through Central Europe to the Black Sea. Rainfall was down to between 50% and 75% of normal levels in Scandinavia and northwestern Europe. Over the western Mediterranean and northern Africa, by contrast, rainfall was much higher than usual, bringing relief to drought-stricken Spain.

In many of these regions, this was a complete reversal of conditions that had prevailed with few interruptions since the mid-1970s¹. Can such marked transitions in European climate be predicted? The study by Rodwell *et al.* on page 320 of this issue² offers renewed hope that they can.

Patterns of temperature and rainfall variability in European winters have long been seen as the fingerprints of the North Atlantic Oscillation (NAO)^{3,4}. The NAO is often defined in terms of the difference in sea-level pressure (the barometric pressure adjusted to the mean sea level) between Iceland and the Azores, and is the tendency of the winter-time mean North Atlantic atmospheric circulation to oscillate between two states: strong and weak. These states differ from one another in the intensity of the regional pressure centres, the Icelandic low and the Azores high (Fig. 1). The switches in sea-level pressure from one state to the other result in changes in mean wind speed and direction, the paths of winter storms, and heat and moisture transport between the Atlantic and the neighbouring continents^{1,3–5}. No wonder, then, that meteorologists have long been trying to predict the state of the NAO ahead of each winter season⁶.

In their study, Rodwell *et al.*² numerically integrate a climate model, consisting of a system of equations that describe the large-scale atmospheric balances of momentum, heat and moisture, with more heuristic schemes that describe small-scale processes such as condensation, cloud formation, precipitation and heat exchange at the land and sea surface. In such modelling work, some variables are set externally: the seasonally changing solar radiation at the top of the atmosphere, the atmospheric carbon-diox-

ide and ozone content, and the sea-surface temperature (SST) and sea-ice distribution. These models are similar to weather prediction models, but unlike the latter are integrated for many years at a time, so that their evolution does not depend on initial conditions. Thus they are not meant to predict the actual evolution of the atmosphere but rather to depict, in a statistical sense, the pattern and strength of atmospheric variability.

In Rodwell and colleagues' study, however, actual SST values over the entire world ocean are used, prescribed to follow their observed monthly states between 1870 and 1997. If atmospheric variability in this model integration is temporally synchronized with observations, the cause is in the ocean.

Over most of the world ocean, variations in SST depend largely on heat exchange with the atmosphere, and so are shaped by atmospheric variability^{7,8}. The ocean itself remains passive, sluggishly accommodating fast-changing fluctuations in atmospheric heat flux to form large-scale SST anomalies. There are only a few examples of an active oceanic role in SST variability, the most

notable being the El Niño phenomenon where large variations in ocean heat transport in the tropical Pacific are brought about by a 'coupled' ocean–atmosphere interaction. Such interaction cannot take place without a strong atmospheric response to changes in SST, and vice versa⁹. Atmospheric response to tropical Pacific SST forms the basis for numerical climate prediction in and around the Pacific Basin: when atmospheric climate models are run with El Niño SST conditions, the models display a realistic basin-wide response¹⁰.

Climate modellers have tried to show that SST anomalies outside the tropics also exert a deterministic influence on the atmosphere, and could therefore also be applied to climate prediction. For this purpose, atmospheric models must be sensitive to the prescribed extratropical SST distribution. However, most attempts to simulate the atmospheric response to extratropical SST anomalies have been frustrating, yielding model-specific results that are inconsistent and generally weak^{7,10}. In what constitutes a breakthrough in this research, Rodwell *et al.* have produced strong evidence for a synchronicity between model-simulated and observed NAO fluctuations, resulting from specifying realistic SST variability in the extratropical North Atlantic Ocean.

Does this bring us closer to predicting the NAO and its effect on Europe's climate? Only up to a point. In the extratropics atmospheric and SST anomalies evolve together, which makes it difficult to specify the SST field ahead of time. This poses a considerable obstacle to a prediction algorithm based on

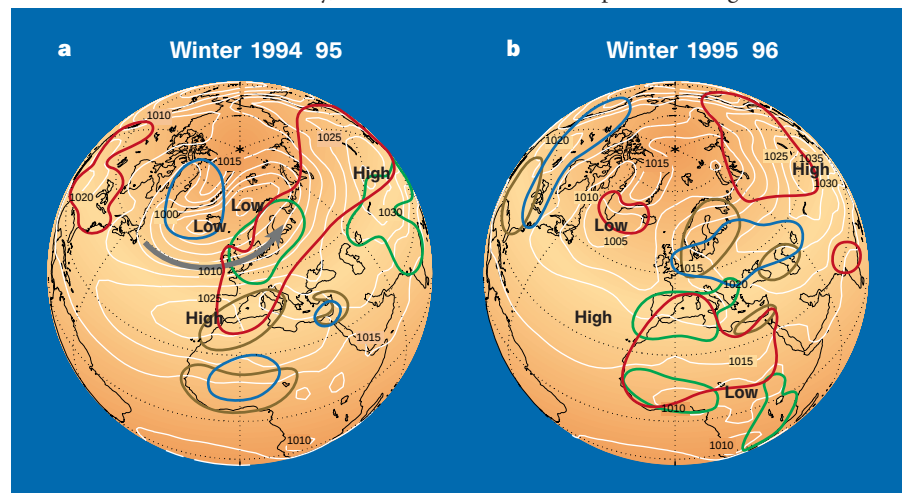


Figure 1 Swings of the North Atlantic Oscillation from one phase to the other affect European climate. a, During the winter of 1994–95, the Icelandic low-pressure and Azores high-pressure regions were both more intense than usual. In this positive phase of the NAO, strong westerly winds (arrow) steered Atlantic winter storms eastward and brought unusually wet conditions to northern Europe and warm temperatures all the way into Siberia. The western Mediterranean was dry and northern Africa cool. b, In the winter of 1995–96 the NAO swung into its negative phase. A weak North Atlantic pressure system led to reduced Atlantic influence over northern Europe, with less rainfall and much colder temperatures. Over the Iberian Peninsula and northern Africa drought gave way to excess rainfall, as winter storms moved freely into the Mediterranean. (Contours depict the seasonal mean sea-level pressure field; regions in red and blue were respectively warmer and colder than normal; regions in green and brown respectively experienced higher and lower rainfall than normal.)

Rodwell and colleagues' study. Advances in ocean modelling may overcome the difficulty, but only in the long term.

Meantime, we need to be able to reproduce Rodwell and co-workers' result with other atmospheric models, and to understand the mechanisms by which the atmosphere 'feels' the SST anomalies. The authors suggest a thermodynamic mechanism in which SST anomalies induce changes in the heat flux from ocean to atmosphere. With this mechanism, a positive SST anomaly would warm the overlying atmosphere through turbulent heat transfer and convection, followed by a change in the circulation. This however implies that, contrary to observations, the atmosphere damps the SST anomalies. As Rodwell *et al.* suggest, such a discrepancy is best addressed with coupled

ocean-atmosphere models where both ocean and atmosphere are allowed to interact. □
Yochanan Kushnir is in the Lamont-Doherty Earth Observatory of Columbia University,
PO Box 1000, Route 9 West, Palisades,
New York 10694-8000, USA.
e-mail: kushnir@ldeo.columbia.edu

1. Halpert, M. S. & Bell, G. D. *Bull. Am. Meteorol. Soc.* **78**, S1-S49 (1997).
2. Rodwell, M. J., Rowell, D. P. & Folland, C. K. *Nature* **398**, 320-323 (1999).
3. Walker, G. T. & Bliss, E. W. *Mem. R. Meteorol. Soc.* **4**, 53-84 (1932).
4. van Loon, H. & Rogers, J. C. *Mon. Weath. Rev.* **106**, 296-310 (1978).
5. Hurrell, J. W. *Science* **269**, 676-679 (1995).
6. Ratchiff, R. A. S. & Murray, R. Q. *J. R. Meteorol. Soc.* **96**, 226-246 (1970).
7. Frankignoul, C. *Rev. Geophys.* **23**, 357-390 (1985).
8. Cayan, D. R. *J. Phys. Oceanogr.* **22**, 859-881 (1992).
9. Philander, S. G. H. *Nature* **302**, 295-301 (1983).
10. Lau, N.-C. *Bull. Am. Meteorol. Soc.* **78**, 21-33 (1997).

Neurobiology

The eyes have it!

Karl Gegenfurtner

The transmission of visual signals from eye to brain involves considerable delays in conduction and processing¹. Because stimuli of varying intensity or colour can cause different delays, it could be difficult to synchronize events from different parts of a visual scene — in particular, our perception of moving stimuli would consistently trail behind their real locations. But the visual system can circumvent such delays by anticipating the path of moving stimuli. Such motion anticipation was assumed to be controlled by high-level motion areas of the visual cortex. Now, very much to our surprise, Berry *et al.*² (page 334 of this issue) report that motion anticipation is already accomplished to a large extent in the retina, by neural circuits that were discovered long ago.

Judging the location of moving objects is important for evading obstacles or predators and for catching prey. These tasks would be almost impossible if the relevant information was delayed. If, for example, we assume a processing delay¹ of about 100 ms, an animal (or a car nowadays) moving at a speed of 40 km per hour would be seen more than one metre behind its actual position. To overcome this potentially lethal problem, evolution has selected mechanisms that anticipate the path of motion.

The existence of mechanisms that compensate for visual delays was uncovered by clever psychophysical experiments, which compared the perceived locations of flashed and moving objects³⁻⁶ (Fig. 1). If presented at the same position, flashed objects are seen to trail moving objects by as much as 80 ms. The common interpretation of this phenomenon was that both types of stimuli — flashed and moving — go through the same delays in the eye, and that the position of the

predictable, moving stimulus is then corrected by movement-selective mechanisms in areas of the brain concerned with analysing visual motion.

In a stunning surprise, Berry *et al.*² now show that motion anticipation not only starts at the retina, the first stage of processing in the visual system, but that it also follows from current models of retinal processing⁷. The basic ingredients are all well studied and common to many stages of processing in the visual system⁸. So how do these ingredients work to produce motion anticipation? The most important part of the process is actually the simplest — namely that retinal ganglion cells pool their inputs over large regions of the visual scene (their receptive fields). In the case of a moving bar, a neuron will start firing as soon as the bar enters its receptive field (Fig. 2). This leads to activity before the bar reaches the centre of

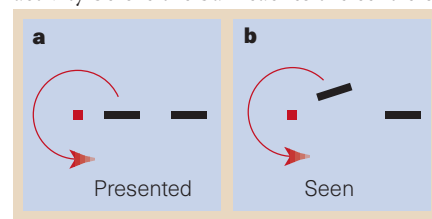


Figure 1 Illustration of motion anticipation. Because of processing delays, our perception of moving stimuli would trail behind their real positions were it not for the development of mechanisms to anticipate the path of moving stimuli. This phenomenon has been demonstrated by psychophysical experiments in which a stationary bar is flashed at the same time as a rotating bar is aligned with it (a). Rather than both bars being perceived at the same position, the moving bar is seen ahead of the stationary bar (b).

the field, yet, because of the transduction processes in the photoreceptors and retinal network, the peak firing rate would still occur some time after the bar has passed the centre.

Of course, it is physically impossible to shift the firing of neurons forward in time. But this is not necessary. The important feature for later processing seems to be the time when the firing rate reaches its peak, and this peak can be shifted forward simply by decreasing the firing rate after a while. That is what the two other ingredients of Berry and colleagues' model do. First, the temporal response of the cell is biphasic and leads to inhibition after a certain time delay. Second, contrast gain control sets the cell's overall sensitivity to stimulation in inverse proportion to retinal contrast⁹. Contrast gain control is a mechanism similar to light adaptation, only it acts on differences in intensity (contrast). So, both of these mechanisms actively inhibit the cell, causing the peak of firing to occur when the leading edge of the bar crosses the centre of the receptive field.

Berry *et al.*² used rabbits and salamanders in their experiments. They show that these 'lower' species clearly do have motion anticipation, meaning that this is an important property that was implemented very early during evolution of the visual system. But are their results valid for primates as well? The retina of lower species contains cells that are selective to the direction of motion, whereas in primates motion processing is done in the cortex. Although Berry and colleagues show that directionally selective cells were not

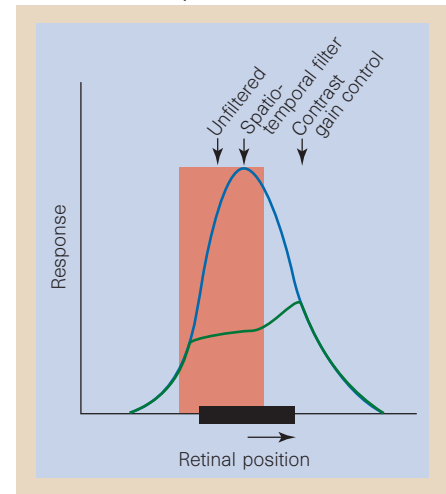


Figure 2 Components of motion anticipation in the retina. The graph represents the 'neural image' of a moving bar at one point in time. The physical position of the bar is shown in black, and its (hypothetical) delayed image in red. Spatio-temporal filtering (blue curve) widens the neural image and shifts the peak firing activity of neurons towards the centre of the bar. Contrast gain control (green curve) shifts the response peak even further ahead towards the leading edge of the bar, as observed by Berry *et al.*².

directly involved in motion anticipation, these cells could, nevertheless, contribute indirectly. I don't think that we need to worry overly about this. The basic mechanisms described by Berry *et al.* are common in many species and at many stages of the primate visual system⁷⁻⁹. This has the advantage that further delays in transmission from the retina to the cortex can be corrected using the same mechanisms.

More importantly, looking back at the human psychophysical literature, there are many indications that motion anticipation occurs in the retina. For example, during sudden, abrupt reversals of motion, the perceived location of the reversal is not overshoot⁵—a finding that argues against cortical extrapolation. Furthermore, whereas most aspects of human visual motion processing are independent of contrast¹⁰, the magnitude of motion anticipation strongly depends on it⁶. And, most significantly, a study by Nijhawan⁴ on colour vision indicates that motion extrapolation occurs in the retina.

Nijhawan flashed a red line on a moving green bar, which should lead a viewer to perceive yellow if inputs from retinal red and green cones are combined. Instead, the red and green were de-coupled, and viewers observed that the flashed red bar trailed the moving green bar. Guided by the firm assumption that motion perception (and, therefore, motion anticipation) occurs in the

cortex, Nijhawan concluded that the synthesis of red and green cone outputs into 'yellow' occurs in the cortex, a conclusion that was regarded with great suspicion by colour scientists¹¹. But the alternative idea—that motion anticipation was done in the retina—seemed even more far-fetched.

Now, Berry *et al.* show that motion anticipation is achieved very well in the retina. Their model calculations show that motion anticipation has nothing to do with motion perception. Instead, it follows in a straightforward way from visual processing mechanisms that were known all along. One wonders what other 'high-level' phenomena could be solved by the low-level retinal circuitry. □

Karl Gegenfurtner is at the Max Planck Institut für Biologische Kybernetik, Spemannstrasse 38, D-72076 Tübingen, Germany.
e-mail: karl@kyb.tuebingen.mpg.de

- Schmolsky, M. T. *et al.* *J. Neurophysiol.* **79**, 3272–3278 (1998).
- Berry, M. J., Brivanlou, I. H., Jordan, T. A. & Meister, M. *Nature* **398**, 334–338 (1999).
- Nijhawan, R. *Nature* **370**, 256–257 (1994).
- Nijhawan, R. *Nature* **386**, 66–69 (1997).
- Whitney, D. & Murakami, I. *Nature Neurosci.* **1**, 656–657 (1998).
- Purushothaman, G., Patel, S. S., Bedell, H. E. & Ogmen, H. *Nature* **396**, 424 (1998).
- Victor, J. D. *J. Physiol.* **386**, 219–246 (1987).
- Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. *Proc. Natl Acad. Sci. USA* **93**, 623–627 (1996).
- Shapley, R. M. & Victor, J. D. *J. Physiol.* **285**, 275–298 (1978).
- Nakayama, K. & Silverman, G. H. *J. Opt. Soc. Am. A* **2**, 267–274 (1985).
- Cavanagh, P. *Nature* **386**, 19–20 (1997).

Seismology

Free oscillations illuminate the mantle

R. Widmer-Schnidrig

Large earthquakes radiate seismic waves that pass through the entire volume of the Earth. Because they are trapped inside the Earth, these waves experience multiple reflections, and interference between waves travelling in opposite directions eventually leads to a standing wave pattern: the 'normal modes' of the Earth (Fig. 1). These 'body' waves are also affected by the medium through which they pass, so they have encoded in them valuable information regarding the distribution of elastic properties and the location of discontinuities.

In two articles published in the *Journal of Geophysical Research*^{1,2}, Resovsky and Ritzwoller analyse frequency spectra of ground motion, from globally distributed seismic observatories, to model the heterogeneous distribution of elastic properties of the Earth's mantle. Free oscillations are most sensitive to the elastic parameters controlling the shear-wave velocity (the velocity of transversely polarized body waves), which is itself

a function of the chemical composition, pressure and temperature of the mantle.

Whether viewed as thermal or chemical anomalies, images of the lateral variation of the speed of high-frequency seismic waves provide some of the most direct evidence for a vigorously convecting, dynamic Earth. However, the style of convection that the mantle is undergoing, and whether it involves a single layer, double layer or something more complex, remains disputed.

In spectra from typically 1–3-day-long seismograms the normal modes show up as discrete peaks, and whereas the locations of the peaks constrain the mantle structure averaged over spherical surfaces, split peaks are evidence of large-scale asphericities and a heterogeneous Earth. Resovsky and Ritzwoller analyse the splitting of 90 normal modes, each yielding independent linear constraints regarding the distribution and size of anomalies in the mantle. Many of the analysed modes are sensitive to structure in the mid-mantle, a region notoriously ill-

sampled by both body and surface waves. Furthermore, the authors fully account for coupling between neighbouring modes, which enables them to obtain constraints for Earth structure of odd spherical harmonic degree. (Simple shapes represented by spherical functions of even and odd harmonic degree are the pumpkin and pear, respectively.) This is the first time that mantle structure of both odd and even spherical harmonic degree have been constrained from normal-mode spectra, providing a more complete picture of the Earth's mantle.

This work has several advantages over previous seismic tomography studies^{3,4}, despite a much lower resolution. By their very nature, normal modes involve the whole volume of the Earth (down to some characteristic depth), and uneven sampling is not an issue. In this respect, normal-mode studies differ markedly from tomographic studies where the structure is only sampled along narrow ray tubes and large volumes remain unprobed by the rays. Another difference is that tomographic studies are good at imaging patterns of heterogeneity, whereas normal-mode studies place tight constraints on the absolute amplitude of these heterogeneities.

The main drawbacks of normal-mode studies are the limited resolution (~2,500 km) and dominant sensitivity to even harmonic-degree structure. Global tomographic studies, by comparison, can resolve structures ~500 km in diameter. Two notable examples^{3,4} feature high-resolution images of individual tectonic slabs subducted into the lower mantle in central America. Now, for the first time, a smoothed-out image of such anomalies seems to be a robust feature of Resovsky and Ritzwoller's models¹ based on normal modes. The fact that these subducted slabs appear as connected features across the transition zone and the mid-mantle is taken by the authors as evidence that the mantle is convecting as a whole. As such, this model is at odds with geochemical data consisting of isotope concentrations in basalt rock samples from mid-ocean ridges and oceanic islands. These data seem to suggest reservoirs of isotopically distinct material in the upper and lower mantle, which in turn favours a layered model of mantle convection without significant mixing across the layers⁵.

How can we make progress on this contradictory view of the mantle? Currently, the seismic models of mid-mantle structure (900–1,800-km depth) put forth by different research groups differ by amounts comparable to the size of the anomalies in the models¹. One promising way to improve this situation is by extending normal-mode analyses to higher frequency modes. The introduction of new methods for the analysis of normal-mode spectra, such as the fully linear,

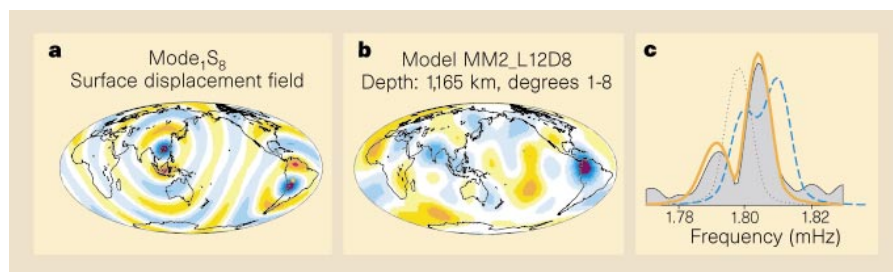


Figure 1 Normal-mode oscillations of the mantle. **a**, A snapshot of the vertical displacement field associated with one normal mode analysed by Resovsky and Ritzwoller^{1,2}. The mode ${}_1S_3$ is sensitive to structure in the lower mantle. The wave field is from the 9 June 1994 Bolivia earthquake (star) calculated for a spherically symmetric Earth model. Red and orange colours indicate up- and downward displacement of the free surface with the maximum displacement reaching 15 μm . **b**, A horizontal cross-section through the favoured model (MM2_L12D8) of the shear-wave velocity perturbations. The depth of this section is 1,165 km, and includes structure up to spherical harmonic degree $l = 8$. Blue regions have increased shear-wave velocity and orange regions have reduced shear-wave speed. **c**, As the structure in **b** is sampled by the wave field in **a**, this wave field gets distorted (not shown here) and the observable spectrum is altered from a simple peak, expected for an Earth with no lateral variation of elastic properties (dotted), to the observed split peak (grey). The splitting produced by modelling the largest asphericity of the Earth, its diurnal rotation and associated hydrostatic flattening (dashed blue), underestimates the observed splitting width and provides a poor fit to the spectrum. A much better fit is provided by the prediction of model MM2_L12D8 (orange). This illustrates a typical spectrum used to infer mantle structure within a normal-mode framework.

autoregressive technique used to estimate the splitting matrix⁶, also holds great potential.

With the global distribution of seismic observatories getting denser and more uniform, the data necessary to constrain structures of shorter wavelength is about to become available. Plus, the overwhelming computational burden of systematically analysing coupled normal modes has now been overcome. This leaves us with the prospect of obtaining enough normal-mode constraints to independently estimate both the heterogeneous velocity and the density distributions in the mantle. This is good news — knowing the density structure (which is invisible to tomographic studies) is crucial to discriminate between styles of

mantle convection, because it is buoyancy stemming from density anomalies that drives mantle convection. □

R. Widmer-Schmidrig is at the Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics, Scripps Institute of Oceanography, University of California San Diego, La Jolla, California 92093-0225, USA.

e-mail: widmer@epicenter.ucsd.edu

1. Resovsky, J. S. & Ritzwoller, M. H. *J. Geophys. Res.* **104**, 993–1014 (1999).
2. Resovsky, J. S. & Ritzwoller, M. H. *J. Geophys. Res.* **103**, 783–810 (1998).
3. Van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. *Nature* **386**, 578–584 (1997).
4. Grand, D. P., van der Hilst, R. D. & Widiyantoro, S. *GSA Today* **7**, 1–7 (1997).
5. Hofmann, A. *Nature* **385**, 219–229 (1997).
6. Gilbert, F., Laske, G. & Masters, G. *EOS Trans. AGU* **79**, 627 (1998).

Oceanography

Phytoplankton death in the sea

David L. Kirchman

It used to be easy to understand how marine phytoplankton are killed off. Grazing by zooplankton and irreversible sinking into the deep ocean were thought to balance phytoplankton growth, except at the beginning of spring blooms when biomass greatly increases. Now, a paper published by Agustí *et al.*¹ in *Limnology and Oceanography* reminds us of a different fate for phytoplankton — these microscopic cells may just die and lyse, without being touched by a grazer.

How phytoplankton die largely determines how other marine organisms live. The classic food chain starts with phytoplankton being eaten by zooplankton which, in turn, are eaten by larger organisms. Running

alongside this food chain is the 'microbial loop', with the production of dissolved organic material (DOM) as its first step. Oceanic organic carbon necessarily starts off in phytoplankton, but DOM also may be produced by zooplankton. Regardless of its source, DOM is taken up by heterotrophic bacteria (those that derive their carbon from organic sources). These bacteria are then eaten by several types of organism, eventually leading back to the classic food chain.

The work of Agustí *et al.*¹, along with other studies on cell lysis, helps to explain previous observations of high bacterial activity in the oceans. Bacterial respiration (oxygen consumption) has been found to

exceed primary production (oxygen production) in oceans poor in plant growth², although this has been questioned³. Less controversial is the observation that bacteria and the rest of the microbial loop consume about half of primary production, implying a high flux through the DOM pool. This 50% estimate is a bit difficult to explain if primary production is first routed through zooplankton to produce DOM. But lysis of phytoplankton cells helps to solve the problem by directly producing DOM for the microbial loop, cutting out the grazer middleman.

The clever method used by Agustí *et al.* to estimate lysis rates of natural phytoplankton assemblages was first proposed by van Boekel *et al.*⁴. These authors followed the release of esterase into sea water. Esterase was picked because it is a cytoplasmic enzyme, its activity can be measured easily using a fluorescent compound (fluorescein diacetate), and algae have more of it than heterotrophic bacteria and grazers. Agustí and colleagues examined several aspects of the esterase method, and provide some support for the last assumption. But more testing may be needed. For example, it is not clear why esterase activity should be higher in phototrophs, which use light energy and inorganic carbon sources, than in heterotrophs. Although this assay may not be perfect, it is difficult to dismiss the results from it.

Brussaard *et al.*⁵ first used the esterase method to estimate that lysis accounts for 75% of phytoplankton death at the end of a spring bloom in the North Sea; lysis is relatively low at other times. Agustí *et al.* have also found incredibly high rates of lysis in surface waters in the northwestern Mediterranean Sea, equivalent to 50% of phytoplankton growth on average. How the growth rates were measured is another story. Although the growth estimates can be questioned, it is unlikely that both studies underestimated phytoplankton growth so substantially that lysis rates seem high only by comparison.

These studies clearly indicate that lysis can be high, but estimates exceeding 50% of phytoplankton growth may raise some eyebrows. For example, there is evidence that, in the Pacific Ocean⁶, phytoplankton growth is roughly matched by grazing, leaving few cells to die by lysis. Brussaard *et al.* also provided evidence that grazing and sinking can balance phytoplankton growth at different times of the year. But their data, now supported by Agustí *et al.*, also indicate that cell lysis can be substantial, especially when grazing and sinking rates are low.

Although the high rates of lysis may be a bit difficult to swallow, there is no doubt that marine algae are lysed by viruses — the main mechanism evoked by Agustí and colleagues to explain their phytoplankton data. Many viruses that attack prokaryotic and eukaryotic

ic marine algae have been isolated, and the total number of viruses in sea water is very high, usually exceeding bacterial and algal abundance. The problem comes in estimating rates of virus-induced lysis. There is circumstantial evidence that viruses can end phytoplankton blooms (viral abundance goes up, number of phytoplankton goes down). Moreover, addition of a viral-sized fraction to sea water can reduce phytoplankton activity⁷. But it has been difficult, using more direct methods, to show that a specific virus can lyse substantial numbers of its unfortunate host in the sea⁸. Perhaps we just lack appropriate methods to estimate viral lysis, as current estimates rely on parameters that are notoriously difficult to measure.

The esterase method is certainly straightforward in comparison, but it may detect phytoplankton dying from causes other than viral attack. Although Agustí *et al.* argue against it, 'sloppy' grazing by zooplankton would break open algal cells before ingestion, releasing esterases and other cell contents. Berges and Falkowski⁹ have suggested another type of cell lysis not considered by Agustí and colleagues. Some algal cells may simply die when stressed by a lack of light or nitrogen, perhaps analogous to apoptotic or programmed cell death in multicellular organisms. The distinction between self-destruction and viral attack is important. Both kill the algal cell, but there's a difference in what happens to the cellular material dur-

ing death. Viruses take over the metabolism of the algal cell to synthesize more viral particles, whereas 'autocatalytic' cell death causes the loss of cellular components owing to increased protease activity⁹. So, the amount and composition of DOM released by these two forms of cell lysis undoubtedly differs.

Predictably, the new study¹ raises more questions than it answers, and highlights the need to look more closely at how phytoplankton die. Compared with the spectacular advances (such as the iron story¹⁰) in understanding phytoplankton growth, much less attention has been paid to phytoplankton death. The fate that awaits phytoplankton in different oceans and times sets the course of food chains and biogeochemical cycles in the sea. □

David L. Kirchman is at the College of Marine Studies, University of Delaware, Lewes, Delaware 19958, USA.

e-mail: kirchman@udel.edu

1. Agustí, S., Satta, M. P., Mura, M. P. & Benavent, E. *Limnol. Oceanogr.* **43**, 1836–1849 (1998).
2. del Giorgio, P. A., Cole, J. J. & Cimleris, A. *Nature* **385**, 148–151 (1997).
3. Williams, P. J. L. *Nature* **394**, 55–57 (1998).
4. van Boekel, W. H. M., Hansen, F. C., Riegman, R. & Bak, R. P. M. *Mar. Ecol. Prog. Ser.* **81**, 269–276 (1992).
5. Brussaard, C. P. D. *et al.* *Mar. Ecol. Prog. Ser.* **123**, 259–271 (1995).
6. Landry, M. R. *et al.* *Limnol. Oceanogr.* **42**, 405–418 (1997).
7. Suttle, C. A., Chan, A. M. & Cottrell, M. T. *Nature* **347**, 467–469 (1990).
8. Garza, D. R. & Suttle, C. A. *Microb. Ecol.* **36**, 281–292 (1998).
9. Berges, J. A. & Falkowski, P. G. *Limnol. Oceanogr.* **43**, 129–135 (1998).
10. Hutchins, D. A. & Bruland, K. W. *Nature* **393**, 561–564 (1998).

Turbulence

The elusive 'ultimate state' of thermal convection

Joël Sommeria

A central dogma in turbulence is that its effects become independent of viscosity when the Reynolds number, which compares flow inertia to viscous damping, is sufficiently large. In thermal convection, this assumption implies that the heat flux becomes proportional to the temperature difference, with an exponent of 1/2, as opposed to the 2/7 usually seen in turbulent convection on the laboratory scale. This 'ultimate' regime (with exponent 1/2) is believed to occur in natural media such as the atmosphere, and evidence for a transition toward this regime has been claimed^{1,2} in recent laboratory experiments. Now, on page 307 of this issue, Glazier *et al.*³ describe experiments at Tohoku University involving mercury at high Reynolds number, with no evidence for this new regime. This raises intriguing questions about the expected 'ultimate state' of thermal convection.

In the absence of a theory for turbulence, dimensional analysis provides several dimensionless numbers for estimating

scaling laws. The Reynolds number is one such dimensionless quantity, defined by $Re = UL/\nu$, for a fluid of kinematic viscosity ν , flowing with speed U past a body of size L . In aerodynamics, it characterizes similarities between flows with the same Reynolds number. Turbulence at very high Reynolds numbers is expected to be controlled by inertial effects (as viscosity passively smoothes out the smallest scales of motion), as seen for flows around an aircraft or car. The drag force on a body becomes independent of viscosity, so it must depend only on the velocity U , the size L and the fluid density ρ . Then dimensional analysis leads to a drag force proportional to $c_x \rho L^2 U^2$ (the only combination with the dimension of a force). The non-dimensional coefficient c_x depends only on the shape, and it can be measured at reduced scale in a wind tunnel providing the inertial regime has been reached. Such an inertial regime is never quite achieved for the flow over a flat surface, such as a straight pipe, in which friction weakly depends on the

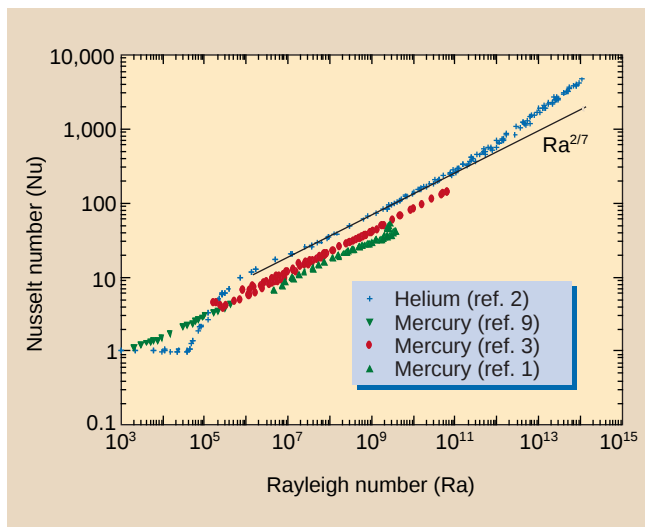
Reynolds number with a logarithmic law. Nonetheless, the drag on a rough surface will become proportional to U^2 at sufficiently high Reynolds number.

In the case of natural convection, velocity is not imposed but is set by buoyancy effects. A central issue is to find a relationship between a temperature difference ΔT applied to the system and the corresponding heat flux Φ . Fundamental studies are often concerned with Rayleigh–Bénard convection of a fluid layer heated from below and cooled from above, but applications are more general. The Rayleigh number, Ra , which is the ratio between thermal driving and viscous dissipative forces, is proportional to the temperature difference, ΔT . The Nusselt number, Nu , is the ratio of the heat flux to the flux that would occur in a purely conducting regime (so it expresses the efficiency of turbulence for enhancing heat transfer). We are therefore seeking a relationship between Nu and Ra . This relationship depends also on the Prandtl number, $Pr = \nu/\kappa$, the ratio of the kinematic viscosity to the heat conductivity κ . Pr varies from about 10^{-2} in liquid metals to 1 for most gases, and more than 1 for liquids such as water and oil.

At very high Ra , the velocity and the transported heat flux are expected to become independent of viscosity and heat conductivity, which is reached in large-scale systems such as the atmosphere. This results in a scaling law, $Nu \propto (RaPr)^{1/2}$, as proposed⁴ in 1962 (indeed, by definition, $Nu \propto \Phi/\kappa$, whereas $(RaPr)^{1/2} \propto 1/\kappa$, so the scaling law eliminates any dependence of Φ on κ and ν).

Experiments in turbulent regimes usually provide power laws such as $Nu \propto Ra^\gamma$ with γ close to 1/3. The 1/3 exponent corresponds to the very simple property that the heat flux is independent of the cell height L (by definition $Nu \propto \Phi L$, whereas $Ra \propto L^3$, so $Nu \propto Ra^{1/3}$ gives a heat flux, Φ , independent of L). This assumption is justified by the observation that the temperature is well mixed and nearly uniform in the bulk flow. Temperature gradients are thus expelled to boundary layers, where velocity is inhibited by wall effects. In the classical viewpoint, turbulence in each boundary layer is generated by local buoyancy effects, resulting in a thermal conductance for each boundary layer, independent of the total fluid thickness. This contrasts with the $Ra^{1/2}$ regime, for which large convective eddies produced in the bulk would fuel shear turbulence near the wall, enhancing heat transfer.

The regime $Nu \propto Ra^{1/3}$ has been verified in experiments at high Prandtl numbers⁵ (using convection of a salt produced by electrolysis instead of heat), but experiments in water ($Pr = 6$) or gas ($Pr \approx 1$) show consistent discrepancy with theory, with an exponent close to 0.29. This exponent was confirmed⁶ with helium gas near its critical point at 5 K, allowing a wide variation of



physical constants, covering a range of Ra , in the same experimental cell. Experiments performed in Chicago⁷ in turn confirmed this result, and produced yet other scaling exponents, involving local temperature fluctuations. A regime of 'hard turbulence' was defined, as opposed to the smoother fluctuations obtained at lower Ra , but this is just the kind of turbulence observed in a saucepan. The Chicago group proposed a model, using scaling estimates for plume velocities and matching with the bulk flow, to provide a good agreement with the measured exponents. The Nu versus Ra exponent was identified as $\gamma = 2/7 = 0.285$.

In principle, this model cannot be extrapolated to the limit of very high Rayleigh numbers. It relies on an assumption that the thermal boundary layer is thinner than the viscous boundary layer, which does not hold in this limit⁸. New regimes, possibly the ultimate turbulence $Ra^{1/2}$, are then expected at higher Ra . Glazier *et al.*³ have clearly reached regimes with thermal boundary layer exceeding the viscous one, but they still find a good fit with the $2/7$ law, without any evidence for changes in the heat flux or turbulence structure. They conclude that the expected ultimate regime may be a chimera.

This work contradicts earlier claims of a transition obtained in mercury¹ (but the limited range of Ra did not allow a firm conclusion). It is also at odds with results in helium gas², but not directly so because the Prandtl number is different (Fig. 1). There is some discrepancy between the new data with helium², and older experiments⁷ (which were more limited in Ra), so the transition may be sensitive to the experimental conditions. A high sensitivity to initial perturbations is observed in transition to turbulence of ordinary boundary layers, for instance in the historical experiment with pipes performed by Reynolds himself, and a similar behaviour could occur for the onset of shear-driven turbulence in the convective boundary layers.

Clearly, further experiments are needed

Figure 1 Dimensionless heat flux (Nu) versus temperature difference (Ra) in Rayleigh–Bénard convection. In mercury ($Pr = 0.025$), data from three experiments agree with the $Nu \propto Ra^{2/7}$ law^{1,3,9}. In helium ($Pr \approx 1$) the data² indicate an increase in Nu , and possibly a trend towards an ultimate regime, $Nu \propto Ra^{1/2}$. Although Ra is smaller in mercury than in helium, the corresponding Reynolds numbers are comparable.

to settle this issue — it is a question of fundamental interest that challenges our understanding of turbulence. It is also of practical use in some engineering problems, for instance when using heat extraction by natural convection for a safer design of nuclear power plants. Extrapolating the scale of a

model may be wrong if transitions to new regimes can occur. Finally, the weather system is a very complex case of natural convection. Here, ground topography or radiative effects probably dominate the boundary layers, so the $Ra^{1/2}$ regime would be expected, but it is difficult to test the absence of well-controlled flow conditions. It will be hard to fully trust a climate prediction as long as the simplest Rayleigh–Bénard convection remains out of reach of turbulence models. □

Joël Sommeria is in the Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69 364 Lyon, France.

e-mail: sommeria@physique.ens-lyon.fr

1. Cioni, S., Ciliberto, S. & Sommeria, J. *J. Fluid Mech.* **335**, 111–140 (1997).
2. Chavanne, X. *et al. Phys. Rev. Lett.* **79**, 3648–3651 (1997).
3. Glazier, J. A., Segawa, T., Naert, A. & Sano, M. *Nature* **398**, 307–310 (1999).
4. Kraichnan, R. H. *Phys. Fluids* **5**, 1374–1389 (1962).
5. Golstein, R. J., Chiang, H. D. & See, D. L. *J. Fluid Mech.* **213**, 111–126 (1990).
6. Threlfall, D. C. *J. Fluid Mech.* **67**, 17–28 (1975).
7. Castaing, B. *et al. J. Fluid Mech.* **204**, 1–29 (1989).
8. Siggia, E. D. *Annu. Rev. Fluid Mech.* **26**, 137–168 (1994).
9. Rossby, H. T. *J. Fluid Mech.* **36**, 309–335 (1969).

Fungal biology

Coming up for air and sporulation

Nicholas J. Talbot

Fungi are familiar to us as mushrooms that we can eat, toadstools that we shouldn't eat, and moulds on food that we have failed to eat. In each case, the visible fungus is composed of thread-like cells called hyphae, which either pack together to form mushroom fruit bodies or build up into a furry mycelium, or mould. In a paper in *Current Biology*¹, Wösten and colleagues reveal that proteins known as hydrophobins constitute the special ingredient that releases these fungal structures from their damp surroundings and enables them to grow up into the air to sporulate.

Fungi spend most of their lives encased in a wet environment such as wood, leaf litter or, in the case of pathogenic fungi, plant or animal tissue. Fungi proliferate by producing extensive hyphal networks that spread in all directions, secreting enzymes to degrade complex nutrients into simple sugars which are taken up to sustain the growing cells. To spread to new territory, however, most fungi need to grow into the air and produce spores. These spores are carried on upwardly projecting aerial hyphae or (in the case of sexual spores) in elaborate fruit bodies such as mushrooms and polypores. Wösten *et al.*¹ have shown that for a fungus to produce aerial structures, it must escape the surface tension of the water that normally surrounds it.

This process involves the action of a remarkable class of fungal proteins called hydrophobins. These are small proteins that

are secreted in abundance by filamentous fungi². They are very diverse in amino-acid sequence but they all have a set of eight cysteine residues and are predominantly hydrophobic in character. In spite of this diversity, even quite different hydrophobins are functionally interchangeable among species, suggesting that they share conserved physical characteristics³. The most extensively studied hydrophobin — and also the one investigated by Wösten *et al.* — is SC3, a hydrophobin produced by the gill-mushroom fungus *Schizophyllum commune* (Fig. 1)^{2,4}. SC3 forms a water-repellent (hydrophobic) outer coating on aerial



Figure 1 Out in the open — the fruit bodies of *S. commune*.

CORBIS/MIKE ZENS

hyphae of *S. commune*, which can be seen under the electron microscope as a layer of interwoven rodlets^{4,5}. Targeted disruption of the SC3 gene produces mutants (Δ SC3) that are unable to form aerial hyphae⁶.

SC3 is secreted as a monomer, but when it encounters an air–water interface or an interface with a hydrophobic surface, it aggregates into a large polymeric complex⁴. This hydrophobin polymer forms a thin layer that is hydrophobic on the side where it is decorated with rodlets, and hydrophilic on the other. This alteration in structure of SC3 occurs in response to changing environmental conditions encountered by the fungus as it grows from water into the air, allowing the fungus to produce hydrophobic aerial hyphae.

But this is only half of the story. SC3 is produced abundantly by *S. commune* even when the fungus is growing in liquids^{2,4} and the importance of this is now clear. In a sim-

ple experiment, Wösten *et al.* showed that the normal surface tension of water, 72 mJ m⁻², can be reduced to as low as 24 mJ m⁻² by adding purified SC3. This makes SC3 the most powerful surface-active protein known and suggests that secretion of hydrophobins at high concentrations dramatically lowers the surface tension, allowing hyphae to escape the liquid and grow into the air. Consistent with this, the surface tension of a liquid culture of *S. commune* falls by a similar amount during growth of the fungus; and when purified SC3 is added to culture medium surrounding a Δ SC3 mutant, the mutant becomes able to put out aerial hyphae.

Aerial hyphae in Δ SC3 mutants that have been treated with exogenous SC3 protein are hydrophilic, indicating that continued secretion of SC3 from aerial hyphae is required for the formation of a hydrophobic surface layer. Thus SC3 has two functions. First, it lowers the surface tension of the water that surrounds the hyphae, allowing them to emerge from the liquid; and second, it coats aerial hyphae with a hydrophobic wall that enables them to grow into the air (Fig. 2) and probably to withstand desiccation.

Since they were first identified in the early 1990s, hydrophobins have been found in many filamentous fungi and may well be ubiquitous². Hydrophobin genes are absent, however, from the genome of the unicellular fungus *Saccharomyces cerevisiae*. The evolution of hydrophobins may therefore have been concurrent with the divergence of the filamentous fungi and, in particular, their colonization of terrestrial ecosystems.

There is still much to be learned about these unusual proteins. Certain fungal species, for example, produce a variety of hydrophobins with diverse biochemical properties⁵, indicating that, in addition to their role in aerial growth, hydrophobins may participate in many other developmental processes. The structural biology of hydrophobins also remains largely unexplored. How, for example, do these small proteins aggregate in response to air–water interfaces, and what conformation do hydrophobin monomers adopt within a rodlet layer? Wösten *et al.* have reached an important milestone in showing how hydrophobins bring about aerial growth in fungi: now we need to determine the molecular basis of their activity. □

Nicholas J. Talbot is in the School of Biological Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter EX4 4QG, UK.
e-mail: N.J.Talbot@exeter.ac.uk

1. Wösten, H. A. B. *et al.* *Curr. Biol.* **9**, 85–88 (1999).
2. Wessels, J. G. H. *Adv. Microbial Physiol.* **38**, 1–45 (1997).
3. Kershaw, M. J., Wakley, G. E. & Talbot, N. J. *EMBO J.* **17**, 3838–3849 (1998).
4. Wösten, H. A. B., Schuren, F. H. J. & Wessels, J. G. H. *Plant Cell* **5**, 1567–1574 (1993).
5. Wösten, H. A. B., Asgeirsdóttir, S. A., Krook, J. H., Drenth, J. H. H. & Wessels, J. G. H. *Eur. J. Cell Biol.* **63**, 122–129 (1994).
6. van Wetter, M. A., Schuren, F. H. J. & Wessels, J. G. H. *FEMS Microbiol. Lett.* **140**, 265–270 (1996).

Daedalus

Press-fit chemistry

Many catalysts provide a passive surface or cavity on which reacting molecules can sit in the right position to react together. Daedalus is now dreaming up a more active type of catalyst.

He recalls the catalytic zeolites, which are riddled with molecular-scale cavities, within which entering molecules can react. Sometimes the cavities actually expand when molecules enter them. So, says Daedalus, imagine an elastic porous catalyst, perhaps a piezoelectric zeolite or microporous silica, expanding and contracting under external control. In its expanded state, it allows reacting molecules to enter its cavities freely. Then it contracts, squashing them together in a way that forces them to react. On the next expansion, the product molecule can escape, to be replaced by new reagents. The cycle could be repeated at megahertz rates.

DREDCO chemists are now trying it. They are combining the skills of the piezoelectric quartz crystal and silica-based catalyst industries to study the effects of setting such catalysts into intense vibration. With luck, the chemical yields of standard test reactions should increase dramatically. Even better, a piezoelectric catalyst could be 'tuned' electronically by setting it into different vibrational modes. A mode which elongated its molecular sites or cavities would give linear molecules; one which flattened them would encourage planar ones. A single catalyst could generate a wide range of products under perfect control.

But the real goal of the project is to find entirely new reactions, yielding hitherto unknown products. Imagine, says Daedalus, a piezoelectric porous catalyst whose cavities can expand to accept three dinitrogen molecules arranged as a hexagon. When such a cavity contracts, the molecules will be squashed together to form the hitherto unknown hexagonal molecule N₆. Being isoelectronic with benzene, it could well be stable. On the next expansion, it would be released, and more dinitrogen taken up. O₆, tetrahedrane (dimerized from acetylene) and many other unknown molecules, could be made in the same way.

N₆ would probably be a powerful high explosive, O₆ a ferocious oxidizing agent, and tetrahedrane an energetic rocket fuel. But they are just the start of a new era of chemical synthesis. A whole bonanza of weird, warped, incredible molecules could pour from the new cornucopia of piezoelectric catalysis.

David Jones

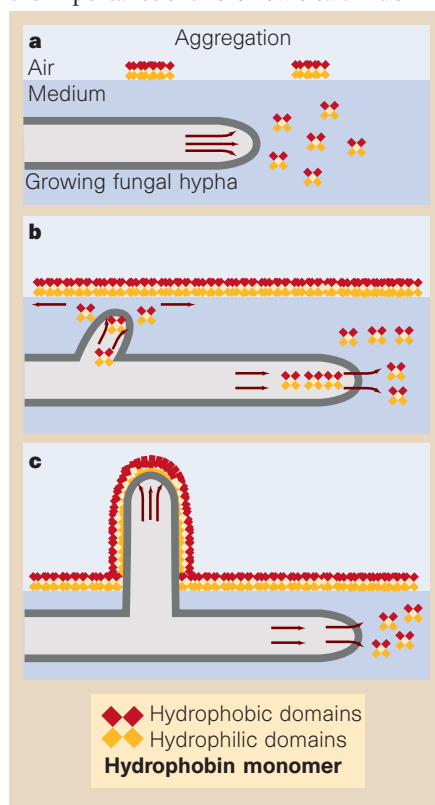


Figure 2 Model for the emergence of aerial hyphae based on the experiments of Wösten *et al.*¹ with the gill mushroom *S. commune*. a, A hypha growing in its normal damp environment, surrounded by a film of water. During growth, SC3 hydrophobin is secreted in abundance. SC3 monomers are depicted as having two hydrophobic domains and two hydrophilic domains² and aggregate at the air–water interface into an amphipathic (with a hydrophobic and a hydrophilic face) layer. b, This causes a dramatic reduction in surface tension, allowing the hypha to escape and grow into the air. Hydrophobin secretion continues. c, The aerial hypha continues to secrete SC3 hydrophobin, which forms a hydrophobic rodlet layer on the outside of the cell wall.

A spelling device for the paralysed

When Jean-Dominique Bauby suffered from a cortico-subcortical stroke that led to complete paralysis with totally intact sensory and cognitive functions, he described his experience in *The Diving-Bell and the Butterfly*¹ as “something like a giant invisible diving-bell holds my whole body prisoner”. This horrifying condition also occurs as a consequence of a progressive neurological disease, amyotrophic lateral sclerosis, which involves progressive degeneration of all the motor neurons of the somatic motor system. These ‘locked-in’ patients ultimately become unable to express themselves and to communicate even their most basic wishes or desires, as they can no longer control their muscles to activate communication devices. We have developed a new means of communication for the completely paralysed that uses slow cortical potentials (SCPs) of the electroencephalogram to drive an electronic spelling device.

The neurophysiological basis and behavioural functions of SCPs have been described², and operant conditioning has been used to bring them under voluntary control. We have shaped the voluntary control of SCPs in two locked-in patients with amyotrophic lateral sclerosis who were able to learn to operate a spelling device by regulating their brain responses. The spelling device drives a cursor on a video screen, allowing the subjects to select letters of the alphabet. Previous interfaces between brains and computers have used different brain responses, such as certain frequency bins or event-related brain potentials^{3,4}, but these have not yet been tested with locked-in patients.

Both subjects suffer from advanced amyotrophic lateral sclerosis and have been artificially respirated and fed for four years. Sustained voluntary control of the musculature was not possible in either patient, so they were unable to use a muscle-driven communication device. They were trained to produce changes voluntarily in their SCPs lasting 2–4 seconds. Each trial consisted of a 2-second baseline and a response period lasting 2–4 seconds, the length of which was adapted to suit the patient. A training day consisted of 6–12 sessions (depending on the condition of the patient), each of which comprised about 70 to 100 trials and lasted about 5–10 minutes.

During the response period, the subjects were required to produce either negativity or positivity greater than a specific criterion amplitude in random order. The baseline and response periods were signalled by two clearly different tones, and whether positivity or negativity was required was indicated

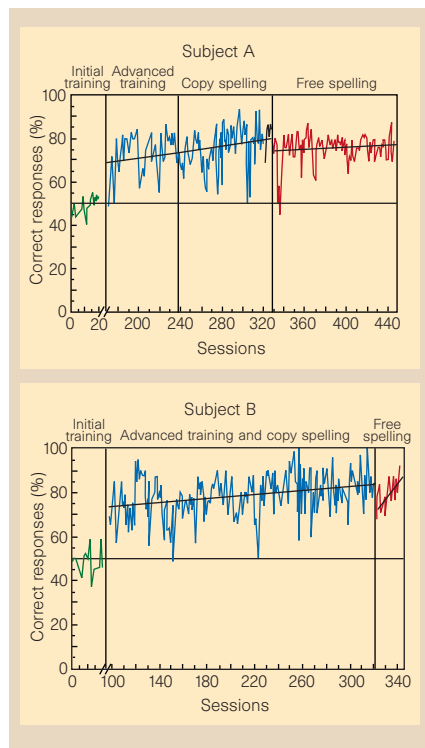


Figure 1 Response accuracy of subjects using the new spelling device. Subject A began with feedback training of SCP amplitude (initial and advanced training; 71.3% correct selections, 75.0% correct rejections, based on ‘go-back’ responses), proceeded to copy spelling (copying of letters and then words; 78.7%, 75.3%) and finally to free spelling (self-selected letters; 66.4%, 82.9%). Subject B began with initial training, then switched to a combination of advanced training (77.5%, 68.8%) and copy spelling (77.5%, 67.6%) and finally to free spelling (86.2%, 73.7%).

by a box being highlighted in either the upper (negativity) or the lower (positivity) half of the screen. Visual feedback of SCPs, which was updated every 64 ms, was provided by a ball that moved towards or away from the box, depending on the direction in which the SCP deviated from baseline. The response criterion was progressively increased from 5 to 8 μ V.

SCPs were extracted from the regular electroencephalogram using a time constant of 8 s and a low-pass filter of 40 Hz. They were recorded from the vertex relative to linked mastoids at a sampling rate of 256 Hz. Vertical eye movements were simultaneously recorded with standard on-line removal of eye-movement artefacts². Using an imagery strategy³, both patients were better able to produce positivity than negativity, so training for negativity was soon discontinued. As soon as a stable performance of at least 75% correct (hitting the goal for at least 500 ms)

LIEBER-HERR-BIRBAUMER-
HOFFENTLICH-KOMMEN-SIE-MICH-BESUCHEN-WENN-DIESER-BRIEF-SIE-ERREICHT-HAT-ICH-DANKE-IHNEN-UND-IHREM-TEAM-UND-BESONDERS-FRAU-KÜBLER-SEHR-HERZLICH-DENN-SIE-ALLE-HABEN-MICH-ZUM-ABC-SCHÜTZEN-GEMACHT-DER-OFT-DIE-RICHTIGEN-BUCHSTABEN-TRIFFT-FRAU-KÜBLER-IST-EINE-MOTIVATIONSKÜNSTLERIN-OHNE-SIE-WÄRE-DIESER-BRIEF-NICHT-ZUSTANDE-GEKOMMEN-ER-MUSS-GEFEIERT-WERDEN-DAZU-MÖCHTE-ICH-SIE-UND-IHR-TEAM-HERZLICH-EINLADEN-DAZU-EINE-GELEGENHEIT-FINDET-SICH-HOFFENTLICH-BALD.
MIT-BESTEN-GRÜßEN-
IHR-HANS-PETER-SALZMANN

Figure 2 The first full message written by subject A.

responses at the 8- μ V level was achieved, the patients began to work with the spelling device. Subject A achieved this level after 327 sessions and subject B achieved it after 288 sessions (Fig. 1).

For the spelling program at level 1, the alphabet was split into two halves (letter banks) which were presented successively at the bottom of the screen for 4.5 seconds (subject A) or 6 seconds (subject B). If the subject selected the letter bank being shown by generating a SCP, it was split into two new halves, and so on until each of the two letter banks had only one letter in it. When one of the two final letters was selected, it was displayed in the top text field of the screen and then the selection began again at level 1. A ‘go back’ function, which appeared as an option when two successive letter banks had not been selected, allowed the speller to go back one step to the previous set of letter banks. If the speller was at level 1, selecting this function erased the last symbol written in the text field, so mistakes could be corrected. This procedure was chosen after extensive pilot work⁶ involving different types of speller.

The accuracy of responses during feedback training, copy spelling and free spelling is shown in Fig. 1 for both patients over the course of 446 (subject A) and 308 sessions (subject B). Two types of error were possible: incorrect rejection was due to the SCP amplitude being too low, and incorrect selection occurred when a high SCP was made in the presence of a wrong letter or row of letters. Both subjects can now write messages. The first full message written entirely by the brain of subject A is shown in Fig. 2.

Our data indicate that patients who lack muscular control can learn to control variations in their SCPs sufficiently accurately to operate an electronic spelling device. Although writing sentences is time consuming — subject A took 16 hours to write the message in Fig. 2, a rate of about 2 characters per minute — it is reliable and precise enough to allow the patient to communicate with his or her environment. These completely paralysed individuals

now have the ability to communicate, a possibility that has not previously existed for such severely affected patients.

N. Birbaumer*†, N. Ghanayim*,
T. Hinterberger*, I. Iversen‡,
B. Kotchoubey*, A. Kübler*,
J. Perelmouter*, E. Taub§, H. Flor¶

*Institute of Medical Psychology and Behavioural
Neurobiology, University of Tübingen,
Gartenstrasse 29, 72074 Tübingen, Germany
e-mail: niels.birbaumer@uni-tuebingen.de

†Department of Psychology, University of Padova,
Via Venezia 8, 35131 Padova, Italy

‡Department of Psychology,
University of North Florida,
Jacksonville, Florida 32224, USA

§Department of Psychology, University of Alabama,
Birmingham, Alabama 35294, USA

¶Department of Psychology, Humboldt-University,
Hausvogteiplatz 5-7, 10117 Berlin, Germany

1. Bauby, J.-D. *The Diving-Bell and The Butterfly* (Fourth Estate, London, 1997).
2. Birbaumer, N., Elbert, T., Canavan, A. G. M. & Rockstroh, B. *Phys. Rev. Lett.* **70**, 1–41 (1990).
3. Vaughan, T. M., Wolpaw, J. R. & Donchin, E. *IEEE Trans. Rehab. Eng.* **4**, 425–430 (1996).
4. Farwell, L. A. & Donchin, E. *Electroencephalogr. Clin. Neurophysiol.* **70**, 510–523 (1988).
5. Roberts, L. E., Birbaumer, N., Rockstroh, B., Lutzenberger, W. & Elbert, T. *Psychophysiology* **26**, 392–403 (1989).
6. Perelmouter, J., Kotchoubey, B., Kübler, A., Taub, E. & Birbaumer, N. *Automedica* (in the press).

Insect antenna as a smoke detector

The larvae of jewel beetles of the genus *Melanophila* (Buprestidae) can develop only in the wood of trees freshly killed by fire¹. To arrange this, the beetles need to approach forest fires from as far as 50 kilometres away^{1,2}. They are the only buprestid beetles known to have paired thoracic pit organs³, which behavioural², ultrastructural⁴ and physiological experiments⁵ have shown to be highly sensitive infrared receptors, useful for detecting forest fires. It has been suggested that *Melanophila* can sense the smoke from fires⁶, but behavioural experiments failed to show that crawling beetles approach smoke sources². We find that the antennae of jewel beetles can detect substances emitted in smoke from burning wood.

We connected freshly excised antennae from *M. acuminata* ($n=5$) to a gas chromatograph equipped with parallel flame ionization and electroantennographic detectors⁷. Volatiles generated by smouldering splint wood from *Pinus sylvestris* were collected on a charcoal trap, chemodesorbed by an organic solvent, and injected into the apparatus. The resulting chromatograms indicated that several components of these volatiles were biologically active (Fig. 1). Most of the volatiles perceived by the antennae are phenolic compounds, derivatives

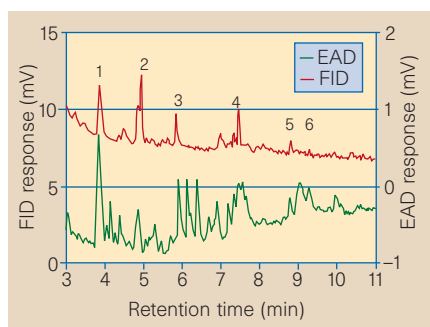


Figure 1 Typical gas chromatograms. Two detectors were used: a flame-ionization detector (FID), indicating any oxidizable compound; and an electroantennographic detector (EAD), using an isolated antenna of *Melanophila acuminata*. The sample contains volatiles released by 1 g of smouldering wood in 1 s, a quantity typically transported over long distances. The FID response indicates quantities of the mixture's components. The EAD response (amplified by a factor of 100) indicates the electrophysiological response of the antenna to: 1, α -pinene; 2, carene; 3, 2-methoxy-phenol (guaiacol); 4, 2-methoxy-4-methyl-phenol (4-methyl-guaiacol); 5, 4-acetyl-guaiacol; 6, 4-formyl-guaiacol. The two peaks only in the EAD trace (retention times of 6.0 to 6.5 min) indicate substances detected by antennae at concentrations below the detection limit of the FID: 10 ng per ml of carrier gas; presumably, they are isomers of compound 3.

of 2-methoxyphenol (guaiacol) eliciting the greatest response.

Methoxylated phenols are released by the incomplete combustion of lignin⁸ and have been identified as atmospheric markers of wood smoke. The chemical structure of the phenolic compounds in smoke is dependent on the species of tree attacked by fire⁹. Because it is particularly sensitive to guaiacol derivatives, *Melanophila* can detect remote forest fires and might even be able to use the pattern of volatiles to identify the species of tree. Our results therefore indicate that the beetles can perceive a fire-damaged *P. sylvestris* tree by olfactory cues.

The beetles' antennae can detect these guaiacol derivatives at concentrations as low as a few parts per billion (p.p.b.) (Fig. 2). We estimate¹⁰ that this sensitivity is sufficient for the beetle to detect a single pine tree 30 cm in diameter that has smouldering bark to a height of 2 m and a bark depth of 1 cm, releasing about 7 g of guaiacol in an hour under light wind conditions (0.3 m s^{-1}), from a distance of more than 1 km.

M. acuminata shows a high absolute ($1.1 \pm 0.8 \text{ pg ml}^{-1}$, $n=14$) and relative sensitivity to guaiacol. For comparison, we also studied the sensitivity to guaiacol of three other beetles. The pulp-feeding forest pest *Phaenops cyanea* attacks weakened trees and sometimes breeds in trees that have previously been damaged by fire. It is closely related to *M. acuminata*, but its attraction

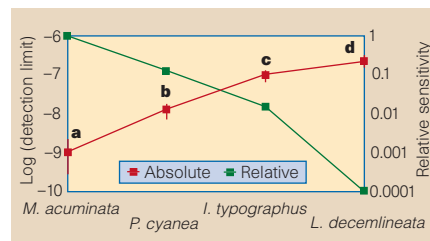


Figure 2 Absolute and relative sensitivities of pyrophilic and non-pyrophilic insects to guaiacol. Absolute detection limits were estimated by fitting logistic curves to dose-response data. The concentrations where confidence intervals (at 0.05) of average control responses did not overlap with confidence belts of fit functions were regarded as the limits of detection. Error bars show standard deviations. Letters indicate significantly differing responses (Mann-Whitney U test, $P<0.05$) to concentrations at detection limits. For insects infesting pine trees, α -pinene was chosen as a reference substance for obtaining relative sensitivities, as it is a typical constituent of host plant odour with a low detection limit. (Z)-3-hexen-1-ol was chosen for the Colorado potato beetle (*L. decemlineata*) as a sensitively detected constituent of its host plant odour.

to fire-damaged trees is less pronounced. Its sensitivity to guaiacol is slightly lower at $15 \pm 7 \text{ pg ml}^{-1}$ ($n=11$). The bark beetle *Ips typographus* is a pulp-feeding forest pest without detectable attraction to fire-damaged trees, but some sensitivity to guaiacols is expected because pulp also contains guaiacol. It shows only a moderate absolute ($120 \pm 65 \text{ pg ml}^{-1}$, $n=9$) and relative sensitivity to this compound. The Colorado potato beetle, *Leptinotarsa decemlineata*, is a non-pyrophilic pest that is not attracted to fire-damaged trees, although it exhibits escape behaviour in response to high concentrations of fire-generated volatiles. It does not show such specialized sensitivity ($450 \pm 160 \text{ pg ml}^{-1}$, $n=24$) to guaiacol derivatives. These values reflect the relative degree of specialization to fire detection.

In the same way that infrared-sensitive pit vipers and boid snakes use chemoreceptors on their tongue to detect volatiles released by their prey, *Melanophila* beetles use chemoreceptors sensitive to the specific volatiles of burning wood. It is unclear how the two sensory systems used by *Melanophila* to identify fire — the thoracic infrared receptors and the antennal olfactory receptors — act together to detect fire and orientate the beetle towards its source.

Understanding how the beetles detect fires could have applications in fire detection devices¹¹ in storehouses and public buildings, for example, as well as in early-warning detection systems for forest fires.

Stefan Schütz*, Bernhard Weissbecker*,
Hans E. Hummel*, Karl-Heinz Apelt†,
Helmut Schmitz‡, Horst Bleckmann‡

*Justus-Liebig-Universität Gießen,

Institut für Phytopathologie und Angewandte

Zoologie, Ludwigstrasse 21 b,
35390 Giessen, Germany
e-mail: stefan.schuetz@agrar.uni-giessen.de
†Landesforstanstalt Eberswalde,
Abteilung Waldschutz, Alfred-Möller-Strasse,
16225 Eberswalde, Germany
‡Rheinische Friedrich-Wilhelms-Universität Bonn,
Institut für Zoologie, Poppelsdorfer Schloss,
53115 Bonn, Germany

1. Linsley, E. G. *J. Econ. Entomol.* **36**, 341–342 (1943).
2. Evans, W. G. *Nature* **202**, 211 (1964).
3. Sloop, K. D. *Univ. Calif. Berkeley Publ. Entomol.* **7**, 1–20 (1937).
4. Vondran, T., Apel, K.-H. & Schmitz, H. *Tiss. Cell* **27**, 645–658 (1995).
5. Schmitz, H., Mürtz, M. & Bleckmann, H. *Nature* **386**, 773–774 (1997).
6. Manee, A. H. *Entomol. News* **24**, 167–171 (1913).
7. Weissbecker, B., Schütz, S., Klein, A. & Hummel, H. E. *Talanta* **44**, 2217–2224 (1997).
8. Sagebiel, J. C. & Seiber, J. N. *Environ. Tox. Chem.* **12**, 813–822 (1993).
9. Edey, L. A. & Richards, G. N. *Environ. Sci. Technol.* **25**, 1133–1137 (1991).
10. Bossert, W. H. & Wilson, E. O. *J. Theor. Biol.* **5**, 443–468 (1963).
11. Schütz, S., Weissbecker, B. & Hummel, H. E. *Biosens. Bioelectron.* **11**, 427–433 (1996).

Rings of single-walled carbon nanotubes

Among the most studied processes of self-organization^{1,2} are the coiling and ring formation of biopolymers such as DNA and proteins. These processes are complex, involving several different types of interaction. We have found that single-walled carbon nanotubes (SWNTs), which are renowned for their extremely high flexural rigidity^{3,4}, can also be induced to organize themselves into rings or coils, with high yields of up to 50%. But unlike coils of biopolymers, in which hydrogen bonding and ionic interactions are usually involved, coils of nanotubes can be stabilized by van der Waals forces alone.

Scanning electron micrographs (Fig. 1a) of SWNTs with an average diameter of 1.4 nm were prepared using laser ablation⁵. The nanotubes were shortened and induced to coil by using an acid treatment with ultrasound. Transmission electron microscope (TEM) images (Fig. 1b) confirm that the rings consist of aligned ropes of SWNT. The size of the rings formed is shown in Fig. 1c.

The structure of rings (tori or coils) can be deduced from several observations. First, long SWNTs are shortened by oxidation, leaving the tube ends functional with carboxylic acid groups^{6,7}, arguing against the formation of a torus involving covalent bonds between carbon atoms. Images obtained by TEM and atomic force microscopy show that the rings do not have a constant thickness and height around their circumference (Fig. 1b), indicating that they may be formed by separate ropes being curled together. The rings can be taken apart

and the ends of the ropes exposed (not shown), so we conclude that they are indeed produced by a coiling process.

Trace quantities (0.01 to 0.04%) of rings have previously been observed⁸ and larger yields have also been claimed⁹. The rings were assumed to be perfect tori⁸, stabilized by covalent bonds between carbon atoms, but our analysis suggests that they were actually coiled SWNTs.

The simplest model of the ring formation process has a SWNT coiling over itself to form a loop. Coiling involves significant strain energy because of the increased curvature, but van der Waals interactions stabilize the tubes.

The critical ring radius, R , for forming thermodynamically stable rings a few micrometres long is small, about 0.03 μm for single tubes or ropes of SWNTs 1.4 nm in diameter. According to our calculations, much lower values of R are energetically allowed than are actually observed, indicating that ring formation may be kinetically controlled. The activation energy, E_A , should be of the order of the strain energy,

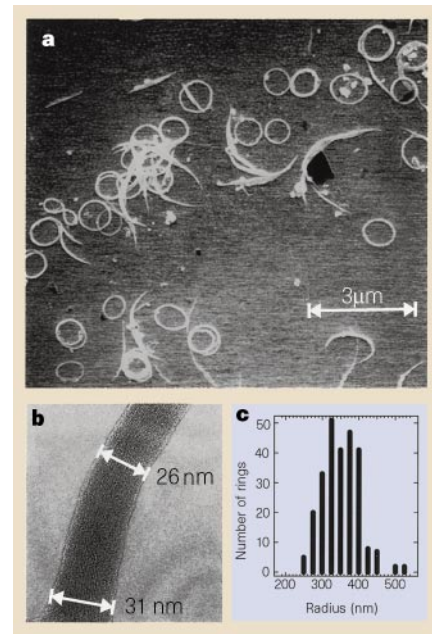


Figure 1 Ring formation in single-walled carbon nanotubes. **a**, Scanning electron micrograph of a SWNT sample dispersed on a hydrogen-passivated silicon substrate, with rings clearly visible. Rings are produced by mixing long SWNTs with a solution of concentrated sulphuric acid and hydrogen peroxide and irradiating them with ultrasound for 1–3 h (40 kHz, 190 W) at 40–50° C, which disperses them and shortens the nanotube ropes⁵. After sonication, the solution is filtered through a 0.2- μm membrane filter, and the residue dried and suspended in 1,2-dichloroethane with a brief period of sonication. A high yield requires shortening the raw nanotubes to a length of 2–4 μm . Yield varies with time of exposure to ultrasound and concentration of peroxide solution. **b**, TEM image of a section of a ring wall (courtesy of L. Gignac). **c**, Histogram showing the distribution of ring radii.

and $\Delta E_A \propto R^{-2}$. In our experiments, the activation energy is provided by ultrasonic irradiation. The most likely mechanism involves the hydrophobic nanotubes acting as nuclei for bubble formation and being bent mechanically at the bubble–liquid interface as a result of the bubbles collapsing during cavitation¹⁰.

Richard Martel, Herbert R. Shea,
Phaedon Avouris
IBM Research Division,
T. J. Watson Research Center,
Yorktown Heights, New York 10598, USA
e-mail: avouris@us.ibm.com

1. Lehn, J.-M. *Supramolecular Chemistry: Concepts and Perspectives* (VCH, Weinheim, 1995).
2. Whitesides, G. M., Mathias, J. P. & Seto, C. T. *Science* **254**, 1312–1319 (1991).
3. Treacy, M. M. J., Ebbesen, T. W. & Gibson, J. M. *Nature* **381**, 678–680 (1996).
4. Wong, E. W., Sheehan, P. E. & Lieber, C. M. *Science* **277**, 1971–1975 (1997).
5. Thess, A. *et al.* *Science* **273**, 483–487 (1996).
6. Liu, J. *et al.* *Science* **280**, 1253–1256 (1998).
7. Wong, S. S., Joselevich, E., Woolley, A. T., Cheung, C. L. & Lieber, C. M. *Nature* **394**, 52–55 (1998).
8. Liu, J. *et al.* *Nature* **385**, 780–781 (1997).
9. McEuen, P. reported in Vossmeier, T. *et al.* *Adv. Mater.* **10**, 351–353 (1998).
10. Suslick, K. S. (ed.) *Ultrasound: Its Chemical, Physical and Biological Effects* (VCH, Weinheim, 1988).

Taxon sampling revisited

Phylogenies that include long, unbranched lineages can be difficult to reconstruct. This is because long-branch taxa (such as rapidly evolving species) may share character states by chance more often than more closely related taxa share derived character states through common ancestry¹. Despite Kim's warning that added taxa can decrease accuracy², some authors have argued that the negative impact of this error, called 'long-branch attraction', is minimized when slowly evolving lineages are included to subdivide long branches^{3–5}. From this they have concluded that increasing the number of species sampled per lineage results in better accuracy than increasing the number of characters per species⁶. We find, using computer simulations, that adding characters can be the more favourable strategy, even for long-branched trees, and that adding slowly evolving taxa to subdivide long branches can reduce accuracy.

An example in which adding characters is a better strategy than adding taxa is shown in Fig. 1a, which compares the performance of alternative strategies for parsimony analyses of data from a 'difficult' tree with four long external branches. This situation could occur, for example, in cases of ancient divergences — such as between fish, amphibians, lizards and mammals — where DNA sequences could be gathered

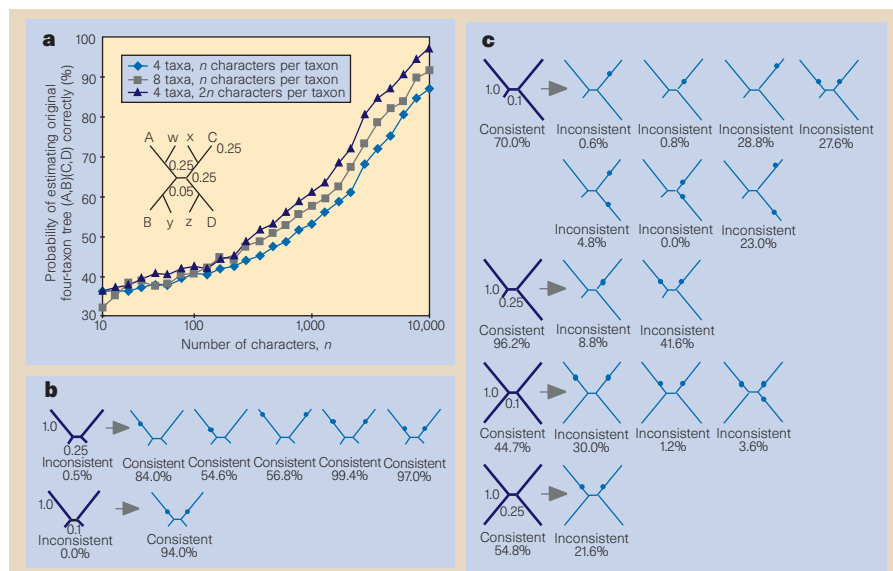


Figure 1 Modelling of phylogenetic strategies. Character data were generated on trees according to a two-state homogeneous, reversible Markov model⁹. Branch lengths are the expected number of changes per character. Trees are unrooted because determining the location of the root requires extrinsic information. PAUP version 4.0 was used for all analyses¹⁰. **a**, For some phylogenies that are difficult to estimate, adding characters is better for accuracy than adding taxa. Parsimony was used to estimate the relationships of A, B, C and D alone or including taxa w, x, y and z. All points are means for 1,000 simulations. **b**, **c**, Effect of long-branch subdivision on the accuracy of phylogeny reconstruction using parsimony, according to consistency and finite data simulation analyses. Bold indicates four-taxon trees for which adding taxa to long branches changed the estimation of relationship of those four taxa from inconsistent to consistent or vice versa. Trees on the right show the 19 (out of 288) cases where adding taxa affected the consistency of the estimation of relationships for the original four-taxon tree. The expected length of each tree is determined from the pattern frequencies predicted under the model and branch lengths. A consistent estimate is achieved if the expected length of the true tree is shorter than that of all others. Percentages refer to the number of times the correct relationships of the original four taxa were recovered in 500 replicate simulations of 5,000 characters. For trees with more than four taxa (all trees where taxa were added), characters were evolved and phylogenetic analyses were run using all taxa, after which 'added' taxa were pruned and resulting four-taxon trees were compared to the true tree. Trees shown are true (not estimated) trees. Dots show where taxa were added. **b**, Cases where long-branch subdivision caused a change from inconsistent estimation to consistency. **c**, Cases where long-branch subdivision caused a change from consistent estimation to inconsistency.

for additional species, assuming they exist, or additional DNA could be sequenced for each species. Adding either characters or taxa usually improves accuracy for original four-taxon relationships. But for the same amount of data, doubling the number of taxa with long-branch subdivision has less impact than doubling the number of characters. The best way to improve accuracy here is to increase the chance of detecting the relatively few changes occurring on the short internal branch, which is better accomplished by adding characters.

Perhaps surprisingly, adding slowly evolving taxa to subdivide long branches can sometimes actually degrade performance. To explore this possibility, we created all of the trees for four taxa composed of branches of two types: long (1.0 expected changes per character) and short (0.1 or 0.25, in separate analyses). We also created all five to eight taxon trees where a slowly evolving taxon (length 0.1) was added to subdivide one to four long branches on the original four-taxon trees. We added one to four taxa (one per branch) to bisect (0.5

from the interior node) or attach basally (0.25 from the node) or distally (0.75 from the node) on the long branch(es). We used parsimony analyses to compare whether the original four-taxon relationships were recovered more or less frequently when taxa were added. We examined consistency (whether the correct tree is obtained with increasing certainty as the amount of character data increases) and performance using a finite number of characters (5,000).

Long-branch subdivision caused changes from inconsistency to consistency (Fig. 1b) as well as from consistency to inconsistency (Fig. 1c). As expected^{3,6}, the problems of inconsistency caused by long-branch attraction in the four-taxon case where two separated lineages are long and other branches are short (the 'Felsenstein zone') can be alleviated by long-branch subdivision. However, no other cases were found where long-branch subdivision caused a change from inconsistent estimation to consistency, and a surprisingly diverse and numerous set of conditions were found where consistently estimated

four-taxon relationships were estimated inconsistently after long-branch subdivision (Fig. 1c).

The explanation for the results shown in Fig. 1c is that long-branch attraction can work for, as well as against, an investigator^{7,8}. Convergent changes may cause spurious attraction between lineages, but they may also help parsimony recover relationships correctly by preventing long branches from being drawn from their proper places by other long branches. Subdividing long branches in such cases can create imbalances of homoplasy that cause the wrong long branches to attract. However, regardless of the topological result, long-branch attraction is always misleading in the sense that undetected changes will cause parsimony to misappropriate changes to internal branches. Other methods, such as maximum-likelihood and corrected-distance methods, explicitly consider undetected changes and are less likely to be affected by taxon-sampling artefacts when the assumed models are appropriate.

Our results show that the commonly recommended strategy of long-branch subdivision should not be applied uncritically. Very dense taxon sampling may yet be crucial for accuracy⁴, and prudent addition of taxa is clearly beneficial in some cases³. However, our examples in which long-branch subdivision is harmful are just as relevant to the choice of strategy as cases in which taxa are added to alleviate long-branch attraction in the Felsenstein zone. One can imagine obtaining the top-left tree in Fig. 1c, noticing the long branches in the recovered tree, suspecting that these branches are spuriously attracting, and then adding taxa to subdivide them as a remedy. In such a case, the situation will appear to be rectified (relationships change), when in fact a problem has been created.

Steven Poe*, David L. Swofford†

*Division of Amphibians and Reptiles, MRC 162, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA and Department of Zoology and Texas Memorial Museum, University of Texas, Austin, Texas 78712, USA

e-mail: stevepoe@mail.utexas.edu

†Laboratory of Molecular Systematics, Smithsonian Institution, Washington DC 20560, USA

1. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).
2. Kim, J. *Syst. Biol.* **45**, 363–374 (1996).
3. Hendy, M. & Penny, D. *Syst. Zool.* **38**, 297–309 (1989).
4. Hillis, D. M. *Nature* **383**, 130 (1996).
5. Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* 2nd edn (eds Hillis, D. M., Moritz, C. & Mable, B. K.) 407–514 (Sinauer, 1996).
6. Graybeal, A. *Syst. Biol.* **47**, 9–17 (1998).
7. Waddell, P. *Statistical Methods of Phylogenetic Analysis*. Thesis, Massey Univ. (1995).
8. Yang, Z. *J. Mol. Evol.* **42**, 294–307 (1996).
9. Cavender, J. A. & Felsenstein, J. *J. Class.* **4**, 57–71 (1987).
10. Swofford, D. L. *PAUP developmental versions* (Sinauer, 1998).

Historical transformations

Master Control Genes in Development and Evolution: The Homeobox Story

by Walter J. Gehring

Yale University Press: 1999. 296 pp. \$35, £25

William McGinnis and Peter A. Lawrence

“It’s a poor sort of memory that only works backwards,” the Queen remarked” — *Through the Looking-Glass* by Lewis Carroll.

The past 20 years have seen embryology transformed into developmental biology: we no longer transplant bits of embryos but run gels, identify genes and try to think up catchy names for them. Walter Gehring has lived through this transition, and he and his co-workers have made many important discoveries on both sides of the divide. The most charismatic of these is the identification of the homeobox, a piece of DNA that encodes a protein motif, a signature written into some of the most illuminating genes ever defined.

In *Master Control Genes* Gehring writes clear and accessible prose, free of the acronyms and stilted jargon of most scientific text. He takes us on a roller-coaster ride, mainly down into the demanding depths of heavy-duty description, occasionally up to the entertaining heights of lyrical reminiscence, giving us a book that is part primer, part autobiography, part gossip. After a Swiss-centric introduction that reviews how information is transferred from genes to proteins, the book introduces its central themes, homeotic genes and the homeobox.

With a sense of wonder, Gehring recounts the history of homeotic mutations that, for example, transform antenna into leg, and then shows how the homeobox became a window through which one could see that humans, bugs and worms are all sisters under the skin, all built and designed by the same genes. This window also showed us that invertebrates and vertebrates share, in part, an immemorial body plan — a grand surprise with evolutionary, even philosophical implications.

The homeobox has also proved to be an aid in picking out some of the special genes concerned in animal design, genes that embody instructions to single cells, directing their differentiation or, more significantly, acting in cell groups, telling them which parts of the body to make. Gehring conveys some of the excitement of the early homeobox era as unifying results were discovered and new models emerged. Indeed, he delights in reliving Ah Hah! moments of discovery, and also reveals a few scars lingering from defunct intellectual disputes. The book then hunkers down to basics, telling in textbook style how embryos are constructed at the cellular level, and how stripes of gene



A homeotic nightmare

This beast is not the work of a mutated “master control gene” but rather of the German artist Thomas Grünfeld. Grünfeld lives and works in Cologne, where, over the past 10 years, he has specialized in taxidermy to produce a whole bestiary of misfits.

In “Misfit (COW)”, the 1997 work pictured here, a bull’s head is apparently seamlessly sewn onto an ostrich’s body. It currently resides in the Saatchi Collection.

expression are generated. He next describes how homeotic proteins may work, and, in a particularly vivid section, how a “master control gene” specifying eye was discovered.

For the general reader, the book brings to life recent discoveries in developmental biology. Historical vignettes garnish the stories: most chapters begin with personal memories from Gehring’s travels that relate to the scientific problems discussed. An interesting section describes how molecular biology was embraced by only a few embryologists like Gehring, who took up cloning with messianic fervour. Today it seems obvious to study both cells and gels, but it cannot have been so in the mid-1970s or more would have done it. Gehring was prescient in this and other decisions and one is at first impressed by his unerring accurate predictions. But after a few of them, as he and his eager subordinates uncover one developmental secret after another, most readers will feel their amazement turning to incredulity. In one passage his predictive prowess takes wing in a time machine, as Gehring assures the reader “I would have predicted [that result]”.

Readers will want to know exactly when, how and by whom the homeobox was discovered. First let’s consider the method: DNA probes made from a homeotic gene were hybridized to gels carrying DNA from many genes. It was found that some probes hybridized to several places on the gel, showing that several genes shared at least a piece of sequence. This piece was the homeobox and turned out to encode about 60 amino acids. Now the history: in the front of the book there is a chronology, starting with *The Origin of Species* and including the homeobox discovery; the former is credited to Charles

Darwin, the latter to Matthew Scott and Gehring himself. Indeed, the discovery was made independently at about the same time in the United States and Switzerland. However, when the key observations were made by Scott and Amy Weiner in Indiana, they were postdoc and graduate student in Thomas Kaufman’s laboratory, and though it was, and is, common practice for lab chiefs to co-author the papers of postdocs and students, Kaufman generously did not insist on it. Kaufman had done fine work, identifying and describing the Antennapedia complex of homeotic genes, upon which Scott and Weiner’s amazing discovery depended, so there is no doubt he deserves considerable credit for preparing the ground in Indiana, just as Gehring himself does in Basel.

Gehring’s account of the discovery in Basel lacks detail and a consistent timeline, and is at odds with the records and recall of other participants. Gehring recalls that he alone saw the significance of his postdoc Rick Garber’s hybridization with Antennapedia DNA. But participants at the meeting (early autumn of 1982) where this was discussed remember that there was an anomalous band in Garber’s gel that could have led to the discovery. However, it was attributed to overloading of the gel, lumped into the ‘uninterpretable results’ category and not followed up. No one besides Gehring can know his thoughts at the time, but if the significance of this band were so obvious, it is a mystery why it was not chased up immediately.

When Bill McGinnis arrived in January 1983, he was not asked by Gehring to explore Antennapedia cross-homology. Instead, he worked on the *lethal giant larvae* gene with

SAATCHI GALLERY LONDON: COURTESY KARSTEN SCHUBERT, LONDON. PHOTO: LOTHAR SCHNEPP

Bernard Mechler, while looking around for other projects. Nothing homeobox-related was done until March 1983, when Michael Levine made the crucial suggestion that an Antennapedia probe be used on a blot with DNA from different species of the fruitfly *Drosophila* (McGinnis was intending to test for sequence conservation in the *lethal giant larvae* gene). This experiment lies at the heart of the discovery of the homeobox — they found that the Antennapedia probe hybridized to multiple bands in both *Drosophila melanogaster* and *D. hydei*, implying common sequences in other genes. As recounted in *The Making of a Fly: The Genetics of Animal Design* by P. A. Lawrence (Blackwell, 1992), Ernst Hafen, Atsushi Kuroiwa, Levine and McGinnis then followed this common sequence to its various sources in the *Drosophila* genome, finding that the sequence mapped to the chromosomal loci of two groups of homeotic genes, the Antennapedia and Bithorax complexes. Then, by hybridizing probes to tissue sections, they found that genes containing the sequence were expressed locally in the antero-posterior axis of embryos. Gehring received DNA from Pierre Spierer that included the Ultrabithorax transcription unit (another homeotic gene) and more cross-hybridization showed that Antennapedia and Ultrabithorax had the same common sequence, which then acquired the homeobox tag.

What about the discovery of homeoboxes in vertebrates? In the book Gehring credits himself and Eddy de Robertis with the decision “in a flare of boldness ... to begin looking for homeobox genes in vertebrates”. But he also points out that there was an earlier “zooblot”, which was done in June 1983. Now, this blot contained DNA from a Noah’s Ark of animals, including vertebrates; it showed up homeobox bands in them all (see *The Making of a Fly*). This discovery made Gehring and De Robertis’ search rather less adventurous than portrayed. It is, of course, a difficult task to write history when nearly all the participants are still alive. Gehring realizes this, stating that the book “represents my personal view and is therefore certainly biased”. Those who have detailed knowledge of some of the events in the book will agree with him on this point.

At the beginning of the most famous autobiographical essay in modern biology, James Watson wrote that he had “never seen Francis Crick in a modest mood”, and Watson goes on to describe events and people with unbridled candour. Gehring’s autobiographical stories provide candour too, but unlike the *The Double Helix*, this book omits the fits and starts, the blind alleys pursued, the struggles with techniques and the endless doubts. It does not illustrate the part that timeliness and luck play in nearly every discovery. In sum, *Master Control Genes* con-



Different perspectives: as Lewis Carroll pointed out, your view depends on where you're standing.

tains a deeply personal view, told in a heroic style, of how a fruitful collision between embryology and molecular genetics of flies and other animals changed how we think about development and evolution. But some of the pictures painted in the book should be viewed through the looking-glass. □

William McGinnis is in the Department of Biology 0349, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0349, USA.

e-mail: wmcginnis@ucsd.edu

Peter A. Lawrence is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

e-mail: pal@mrc-lmb.cam.ac.uk

Painting a picture of development

The Art of Genes: How Organisms Make Themselves

by Enrico Coen

Oxford University Press: 1999. 392 pp. £20

John Maynard Smith

“Over the past twenty years there has been a revolution in biology: for the first time we have begun to understand how organisms make themselves.” These are the opening words of the preface to Enrico Coen’s book, and I think they are true. In *The Art of Genes*, Coen tries to explain to readers with no special knowledge of biology or development what is happening in the world of genetics.

He does so by means of a protracted analogy with artistic creation; but this is not an attempt to obfuscate or mystify. Coen has himself contributed to the revolution that is under way and he is a mechanist at heart, with a passion for communicating. He uses

the artistic analogy to emphasize the difference between the process of development and the fabrication of a machine by following a plan or blueprint. Before a painter starts painting a picture, he does not have an image in his mind of the completed picture. Instead, he reacts continuously to the pattern of paint already on the canvas.

Coen introduces the analogy by describing Curt Stern’s 1956 study of the effects of the *scute* gene on bristle patterns in the fruitfly *Drosophila*. (One of the book’s great merits is Coen’s sense of the history of his subject, from debates between preformation and epigenesis to the premolecular era of developmental genetics.) Stern studied flies that were mosaics of genetically *scute* and normal tissue. He explained the resulting bristle patterns as arising from an interaction between what he called a ‘prepattern’ (corresponding roughly to what Lewis Wolpert later called positional information) and the competence of cells locally to respond to it. He found that the prepattern in *scute* flies was unchanged from the normal but cell competence to respond was impaired.

I remember being both excited and disappointed by Stern’s paper — excited because it showed how genetics could be used to analyse development, but disappointed because the prepattern was unchanged. After all, it is the pattern, not the response, that is the exciting thing we want to understand. There is no longer any need to feel disappointed. Genetic analysis is also revealing how prepatterns are laid down, as Coen illustrates by describing, in some detail, the development of segmentation in *Drosophila* and of flower morphology.

Following the artistic analogy, prepattern is described in terms of regions of differing colour, eliciting different responses. Coen gives a detailed account of early *Drosophila* development along these lines, including a discussion of how colours spread, how some responses depend on the presence of two contrasting colours, and so on. At least the analogy helps him to avoid the clumsy terminology of *Drosophila* development, with its *bicoid* and *hedgehog*, *Notch* and *fushi-tarazu*, and to intersperse the geometric diagrams with pictures by Magritte, Escher, Picasso and Heath-Robinson.

I am delighted that one of the leading practitioners in the field should have taken time off from the lab to tell the rest of us what is going on. How far has the attempt succeeded? It does avoid any assumption of previous knowledge: there is nothing here you cannot understand if you want to. But it is not an easy read. If you want a book entitled “A Brief History of Development”, to browse through for an hour, learning enough of the vocabulary to keep your end up in cocktail conversation, this is not for you.

But there is an alternative reader for whom the book is ideal. When I have

attempted a book at this level, I have had an ideal reader in mind — myself aged 16, largely ignorant but reasonably intelligent, and willing to make a serious effort if the prize in understanding seemed worth it. I would have loved this book at 16, and so should anyone — aged 16 to 60 — who really wants to understand development. □

John Maynard Smith is at the School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK.

Surfing a chaotic sea

Physics of Chaos in Hamiltonian Systems

by G. M. Zaslavsky

World Scientific: 1998. 268 pp. £30, \$44

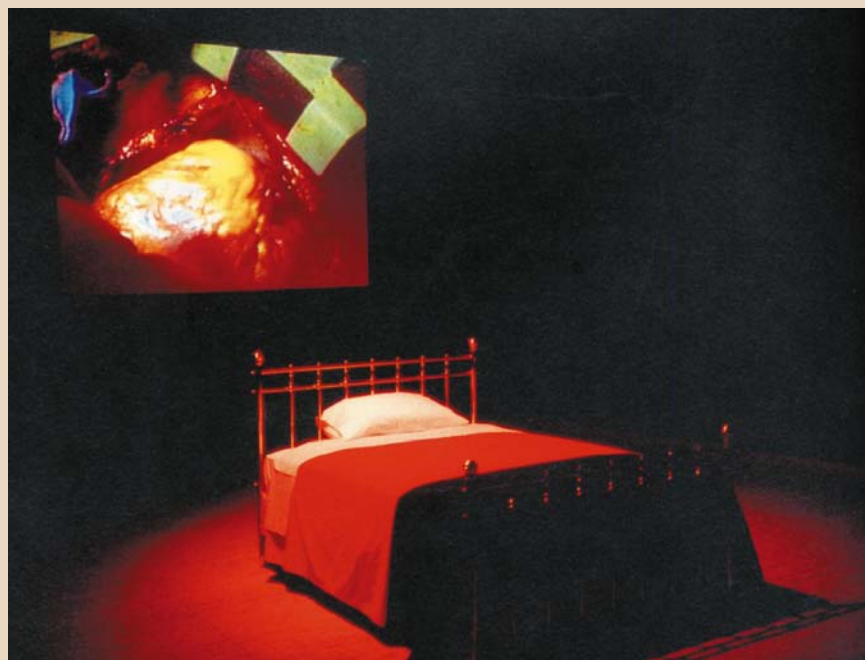
J. D. Meiss

Remarkably, the basic models of motion in physics — ranging from Newton's law of gravitation to modern string theory — can all be formulated as Hamiltonian dynamical systems. Here there is a striking pairing between position coordinates (configuration variables) and their rates of change (momentum variables). Each such pair of variables constitutes a 'degree of freedom' of the system. Hamiltonian dynamics normally conserves energy, but this need not be the case; for example, a frictionless pendulum with an oscillating support is a nonconservative Hamiltonian system. Time dependence is considered a 'half' degree of freedom. The goal of *Physics of Chaos* is to explore the structures and motion of systems with one and a half degrees of freedom.

Hamiltonian dynamics can show extraordinary complexity when the number of degrees of freedom exceeds one. This complexity gives rise to 'fractal' images that are just as visually engaging as the famed Mandelbrot set, and its study is vital for understanding everything from chemical reaction rates to the stability of the Solar System.

Computer studies show that the motion consists of islands of stability in a sea of chaos. These islands can even mimic quasicrystalline patterns. Some of our understanding of such systems can be traced back to Poincaré, but many questions remain. The major one addressed here is: can statistical techniques be applied to such chaotic motion? While these methods apply to 'uniformly' chaotic systems, the ubiquitous islands of stability lead to long correlations that contravene the use of standard statistical (for example, diffusive) techniques. The theory of transport in such systems relies on the geometry of the stable manifolds and of the 'cantori' (structures that act as nearly impermeable barriers) that permeate the chaotic sea; this has been carefully treated by Stephen Wiggins in *Chaotic Transport in Dynamical Systems* (Springer, 1991).

Rhythms of the human heart



A human heartbeat forms the background to Bill Viola's video installation, "Science of the Heart". His work — showing at Frankfurt's Museum für Moderne Kunst until April, in San

Francisco from June and in Chicago from October till January 2000 — is also displayed in the book *Bill Viola* by the Whitney Museum of American Art (Flammarion, \$60, £39.95).

George Zaslavsky develops 'fractional kinetics' in an attempt to give a smoothed, but nondiffusive, description. This phenomenological description captures some aspects of the stickiness of islands, but I believe its mathematical justification remains elusive. Perhaps that is an excellent reason to read this book. □

J. D. Meiss is in the Department of Applied Mathematics, University of Colorado, Boulder, Colorado 80309, USA.

A bloody business

Blood: An Epic History of Medicine and Commerce

by Douglas Starr

Knopf: 1998. 440 pp. \$27.50, £20

Fred S. Rosen

While the value of a barrel of crude oil has fallen to \$12, an equivalent amount of blood is now worth around \$60,000. How this came about is lucidly recounted by Douglas Starr in this history of blood transfusion. The book has received broad critical acclaim in the lay press and is indeed accessible to a wide-ranging audience.

From its murky beginnings in the eighteenth century, blood transfusion became a practical reality in this century through the efforts of many unsung heroes. In 1908, Alexis Carrel, working at the Rockefeller Institute in New York, successfully transfused a profoundly anaemic newborn baby

by connecting her physician father's radial artery to the infant's popliteal vein with a cannula. Soon after, Richard Lewisohn at the Mount Sinai Hospital in New York showed that sodium citrate could be used to prevent blood coagulation without harming the blood or its recipient. By this time, Karl Landsteiner had discovered the ABO blood-group system and two of his students, Philip Levine and Alexander Weiner, subsequently discovered the rhesus (Rh) blood groups. The First World War dramatized the need for blood and blood products to combat shock and other complications of the wounded. The elements were in place for the birth of the blood-banking industry.

The Western democracies and the USSR made great strides during the Second World War in delivering dried plasma and albumin to the battlefield, as fleetingly portrayed in the opening scenes of the recent movie *Saving Private Ryan*. The Germans were bogged down in attempts to preserve the racial purity of the blood supply, a folly matched only by the Americans' rules to segregate the blood of black Americans. It is ironic that a distinguished black American physician directed the first large-scale blood collection in the United States at the beginning of the Second World War — Charles Drew directed the Plasma for Britain campaign.

The outstanding work of Edwin J. Cohn at Harvard Medical School in fractionating plasma into its useful components is described with all the details of showmanship that Cohn conjured up.

But then the darker side of this history reveals itself in the emergence of tainted blood drawn from donors in search of cash. A dismaying chapter in this history involves blood collection for money in Central America and other underdeveloped countries from people in desperate financial need. The problems of hepatitis transmission were soon overshadowed by the contamination of the blood supply by the human immunodeficiency virus (HIV).

Then there is the frightening tale of the deadly triumph of the free market. When the Community Blood Bank of Kansas City, Missouri, became the favoured source of blood for physicians because a local commercial blood-donor centre was collecting contaminated blood, the United States Federal Trade Commission accused the physicians of illegal conspiracy to restrain free trade! A fine of \$5,000 a day was imposed until blood was purchased from the contaminated commercial source.

The last part of the book looks at the now twice-told tale of HIV contamination of anti-haemophilic factor, and the bitter resentments and ugly legal squabbles that ensued. The former French prime minister Laurent Fabius was acquitted only a few weeks ago — to the dismay of many, in what Starr calls cathartic national breast-beating.

Many significant contributors to blood banking do not appear in this book. There is no mention of Allen Latham, who at Cohn's instigation invented the blood-cell separator, the so-called Latham bowl. The monumental struggle of Harvard surgeon Carl Walter with the US Food and Drug Administration to get blood out of breakable glass bottles and into malleable plastic bags is given only brief mention. And the late Wallace Coulter, who standardized blood counting worldwide with his Coulter counter,

invented in the basement of a Chicago boarding-house using the Cellophane wrapper from a pack of Lucky Strike cigarettes during the Second World War, gets no space.

There are many more books to be written about all this. And the blood supply is once again safe — until the next unknown blood-borne virus creeps up on us. □

Fred S. Rosen is in the Department of Pediatrics, Harvard Medical School, Longwood Avenue, Boston, Massachusetts 02115, USA.

Storm-chaser reaps the whirlwind

Tornado Alley: Monster Storms of the Great Plains

by Howard B. Bluestein

Oxford University Press: 1998. 170 pp. \$35, £17.99

Roger A. Pielke Sr

Howie Bluestein is the man who made 'storm-chasing' scientifically acceptable. A pioneer in the field observation and measurement of tornadoes and tornadic thunderstorms, he blends history and science with his personal experiences in a book that will be enjoyed by anyone who is interested in these intense atmospheric whirlwinds. It should also help clear up any misunderstandings that resulted from seeing the recent film *Twister*!

Though the text is very detailed in places, non-specialists can safely skim the technical sections without losing the point. The use of degrees Fahrenheit and feet may be annoying to some readers, but the numbers can easily be converted to metric units.

The book analyses other people's experiences of tornadoes as well as drawing on

Bluestein's personal perspective. It includes a useful discussion of the Fujita scale of intensity, which was introduced by Ted Fujita of the University of Chicago and has now become accepted by the public.

Bluestein looks at the use of computers to simulate tornadic thunderstorms: the correspondence between computer modelling and the actual behaviour of thunderstorms is remarkable. Unfortunately, he does not mention some important recent work by Louie Grasso of the US National Oceanic and Atmospheric Administration and Cathy Finley of the University of Northern Colorado. These two have, separately, simulated the development of actual tornadoes — Finley has even simulated the development and dissipation of their suction vortices. Bluestein also overlooks the modelling work of researchers such as Bob Walko while at the University of Oklahoma, and Conrad Ziegler at the National Severe Storms Laboratory.

Because of the major role he has played in field campaigns studying 'Tornado Alley' in the Great Plains of the United States, Bluestein is often seen in television shows on this now-popular subject. In fact, tornadoes have a long history on our screens. Bluestein's TOTO — the Totally Totable Observatory, a canister packed with meteorological instruments and used in some of his earlier field campaigns — was named after Dorothy's dog in *The Wizard of Oz*. (*Twister* was based on attempts to insert these instruments into the path of tornadoes.) The discussion of TOTO in this book is a pleasure to read.

Readers interested in an overview of measurement systems will appreciate Bluestein's explanations of facilities spanning the size range from helicopters and remotely piloted vehicles to the latest model TOTO II. Portable rawinsondes and Doppler radars, Doppler lidars, millimetre-wavelength radars and airborne Doppler radars — each of these offers a unique way to monitor tornadoes.

What next? Bluestein's speculation on the future of tornado research, as one of the scientific leaders in this field, makes important reading. For his part, he is likely to continue to play an active role in their investigation.

For further reading, the book lists relevant websites and other sources of tornado information. The outstanding set of videos prepared by the Tornado Project in St Johnsbury, Vermont, should be added to the list (<http://www.tornadoproject.com>).

In conclusion, this book by an internationally recognized expert in his subject is an enjoyable read. The scientific understanding of tornadoes is increasing all the time, and Bluestein masterfully weaves his experiences together with his insights into this constantly evolving field of research. □

Roger A. Pielke Sr is in the Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado 80523, USA.



Threatening clouds: a tornado in its mature phase looms over the town of Canadian, Texas, in 1986.

Josephson-junction qubits with controlled couplings

Yuriy Makhlin^{*†}, Gerd Schön^{*} & Alexander Shnirman[‡]

^{*} Institut für Theoretische Festkörperphysik, Universität Karlsruhe, D-76128 Karlsruhe, Germany

[†] Landau Institute for Theoretical Physics, Kosygin Street 2, 117940 Moscow, Russia

[‡] Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Quantum computers, if available, could perform certain tasks much more efficiently than classical computers by exploiting different physical principles^{1–3}. A quantum computer would be comprised of coupled, two-state quantum systems or qubits, whose coherent time evolution must be controlled in a computation. Experimentally, trapped ions^{4,5}, nuclear magnetic resonance^{6–8} in molecules, and quantum optical systems⁹ have been investigated for embodying quantum computation. But solid-state implementations^{10–14} would be more practical, particularly nanometre-scale electronic devices: these could be easily embedded in electronic circuitry and scaled up to provide the large numbers of qubits required for useful computations. Here we present a proposal for solid-state qubits that utilizes controllable, low-capacitance Josephson junctions. The design exploits coherent tunnelling of Cooper pairs in the superconducting state, while employing the control mechanisms of single-charge devices: single- and two-bit operations can be controlled by gate voltages. The advantages of using tunable Josephson couplings include the simplification of the operation and the reduction of errors associated with permanent couplings.

Two versions of Josephson-junction qubits are shown in Fig. 1. The simpler one (Fig. 1a), proposed earlier¹⁰, consists of a superconducting electron box, that is, a low-capacitance island coupled via a Josephson tunnel junction to a lead. The Coulomb interaction (charging energy) restricts the number, n , of Cooper-pair charges, $Q = 2ne$ (where e is the charge on an electron), on the island. If biased near a degeneracy point the system constitutes a qubit with two states differing by one Cooper-pair charge. Quantum logic operations can be performed by switching the gate voltage. Before describing the systems in detail we will first present an ideal model. This puts in perspective the possibilities and drawbacks of the simple design, as well as the advantages of the new design with Josephson coupling controlled by a superconducting quantum interference device (SQUID; Fig. 1b). These are, first, during idle periods between operations the energy splitting between logical states is tuned to zero, thus avoiding an undesired phase evolution. With this drawback of most proposals overcome, the requirement on the precision of time control is substantially reduced; second, the 2-bit couplings can be switched on and off avoiding errors associated with permanent couplings.

To realize a quantum computer we search for a system with the following ‘ideal’ model hamiltonian:

$$\hat{H} = - \sum_{i=1}^N [H_z^i(t)\hat{\sigma}_z^i + H_x^i(t)\hat{\sigma}_x^i] + \sum_{i \neq j} J^{ij}(t)\hat{\sigma}_+^i \hat{\sigma}_-^j \quad (1)$$

A spin notation is used for the qubits with Pauli matrices $\hat{\sigma}_z$, $\hat{\sigma}_x$, $\hat{\sigma}_\pm = \frac{1}{2}(\hat{\sigma}_x \pm i\hat{\sigma}_y)$. Ideally, each energy $H_z^i(t)$, $H_x^i(t)$ and the (real symmetric) couplings $J^{ij}(t)$ can be switched separately for controlled times between zero and finite values. We assume that H_z^i is the largest energy, suggesting the choice of basis states $| \uparrow \rangle$ and $| \downarrow \rangle$ aligned along the z -axis. Residual inelastic interactions (which destroy the coherence), and the measurement device (when turned on) should be accounted for by extra terms $\hat{H}_{\text{res}}$ and $\hat{H}_{\text{meas}}(t)$, respectively.

Quantum computation requires four elementary steps.

(1) The system has to be prepared in a well defined initial state. For this we turn on at low temperature all $H_z^i \gg k_B T$, while $H_x^i = J^{ij} = 0$. After sufficient time the residual interaction $\hat{H}_{\text{res}}$ relaxes all spins to the ground state, $| \uparrow \uparrow \dots \rangle$. Then $H_z^i(t)$ is set back to zero.

(2) Single-bit operations (gates) have to be performed. They are controlled by turning on one of the fields. If H_x^i is switched on, the spin i evolves according to the unitary transformation $U_{1b}^i(\tau) = \exp(iH_x^i \tau \hat{\sigma}_x^i / \hbar)$. Depending on the time span τ , a π - or $\pi/2$ -rotation is performed, producing a spin flip or an equal-weight superposition of spin states. Switching on one H_z^i produces another needed operation: a phase shift between $| \uparrow \rangle$ and $| \downarrow \rangle$. Back in the idle state, where $\hat{H} = 0$, the relative phase shift of the states does not evolve further.

(3) A two-bit operation on qubits i and j is achieved by turning on the corresponding J^{ij} . In the basis $| \uparrow \uparrow \rangle, | \downarrow \downarrow \rangle$, the result is described by $U_{2b}^{ij}(\tau) = \begin{pmatrix} \cos \alpha & i \sin \alpha \\ i \sin \alpha & \cos \alpha \end{pmatrix}$, with $\alpha = J^{ij} \tau / \hbar$, while the states $| \uparrow \downarrow \rangle, | \downarrow \uparrow \rangle$ are not affected. For $\alpha = \pi/2$ the result is a spin-swap operation, while $\alpha = \pi/4$ yields a ‘square-root swap’. The latter transforms the state $| \uparrow \downarrow \rangle$ into the entangled state $(| \uparrow \downarrow \rangle + i | \downarrow \uparrow \rangle) / \sqrt{2}$. The combination with single-bit operations allows us to perform the ‘controlled-not’ gate; in fact, they provide a universal set, sufficient for all logic gates of quantum computations¹⁵.

(4) The final state has to be read out, which constitutes a quantum measurement process¹⁶.

Searching for nanometre-scale electronic realizations of qubits, one might consider normal-metal single-electron devices. But they are ruled out, because in the normal state different tunnelling processes are incoherent. Ultrasmall quantum dots with discrete levels or spin degrees of freedom in nanostructured circuits^{13,14} are candidates, but are difficult to fabricate in a controlled way. More

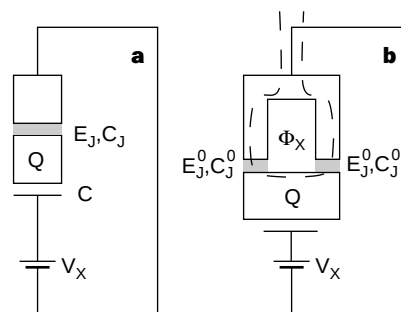


Figure 1 Josephson junction qubits. **a**, A simple realization of a qubit is provided by the superconducting electron box. A superconducting metallic island is coupled by a Josephson tunnel barrier (with capacitance C_J and Josephson coupling energy E_J ; grey area) to a superconducting lead and through a gate capacitor C to a voltage source. The important degree of freedom is the Cooper-pair charge $Q = 2ne$ on the island. **b**, The improved design of the qubit. The island is coupled to the circuit via two Josephson junctions with parameters C_J^0 and E_J^0 . This d.c.-SQUID can be tuned by the external flux Φ_X which is controlled by the current through the inductor loop (dashed line). If the self-inductance L_Φ of the SQUID is low, $\Phi_X^0 / L_\Phi \gg 4\pi^2 E_J^0, e^2 / C_J^0$, fluctuations of the flux from Φ_X are weak. Furthermore, if the frequency of flux oscillations is high, $\hbar \omega_\Phi = \hbar (L_\Phi C_J^0 / 2)^{-1/2} \gg E_J^0, E_{\text{ch}}, k_B T$, the Φ -degree of freedom is in the ground state. In this case, the set-up allows switching the effective Josephson coupling to zero. ($E_J = 0$ requires the Josephson energies of two junctions in the loop to be equal. This has been reached with a precision of 1% in quantum tunnelling experiments¹⁷. Even with this precision, taking into account¹⁰ the finite value of E_J one can perform a large number of logical gates. On the other hand, by replacing one junction in **b** by another SQUID, one can tune the Josephson couplings to be equal.) The effective junction capacitance is $C_J = 2C_J^0$.

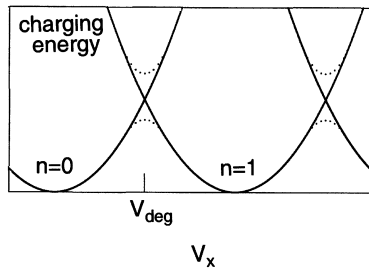


Figure 2 Spectrum of a superconducting electron box. The charging energy, $(Q - CV_x)^2/(2(C + C_j))$, of the superconducting electron box is shown (solid lines) as a function of the applied gate voltage V_x for different numbers n of extra Cooper pairs on the island. Near degeneracy points, the weaker Josephson coupling energy mixes the charge states and modifies the energy of the eigenstates (dotted line). In this regime, the system effectively reduces to a 2-state quantum system.

promising are systems built from Josephson junctions, where the coherence of the superconducting state can be exploited. Quantum extension of elements based on single-flux logic have been considered (ref. 17, and J. E. Mooij, personal communication). Encouraged by successful experiments that demonstrated the superposition of charge states^{12,18,19}, we suggest here the use of superconducting electron boxes with low-capacitance Josephson junctions as qubits.

In the system of Fig. 1a Cooper pairs tunnel coherently, while Coulomb blockade effects allow the control of the charge. The relevant conjugate variables are the phase difference γ across the junction and the charge $Q = 2ne$ on the island. If quasiparticle tunnelling is suppressed by the superconducting gap and only 'even-parity' states are involved²⁰, the circuit dynamics is governed by the hamiltonian:

$$\hat{H} = \frac{(Q - CV_x)^2}{2(C + C_j)} - E_j \cos \gamma; \quad Q = \frac{\hbar}{i} \frac{\partial}{\partial (\hbar\gamma/2e)} \quad (2)$$

For the junctions considered, the charging energy with scale $E_C \equiv e^2/2(C + C_j)$ dominates over the Josephson coupling E_j . It is plotted in Fig. 2 as a function of the external voltage V_x for different n . In equilibrium at $k_B T \ll E_C$, the system is in the state corresponding to the lowest parabola. But, near the voltages $V_{deg} = (2n + 1)e/C$, the states n and $n + 1$ are near-degenerate, and E_j mixes them strongly. Here, in the basis of charge states $|1\rangle = |n\rangle$ and $|1\rangle = |n + 1\rangle$, the hamiltonian reduces to a two-state model

$$\hat{H} = E_{ch}(V_x)\hat{\sigma}_z - \frac{1}{2}E_j\hat{\sigma}_x \quad (3)$$

where $E_{ch}(V_x) = e(V_x - V_{deg})C_{qb}/C_j$, and the capacitance of the qubit in the circuit is $C_{qb}^{-1} = C_j^{-1} + C^{-1}$.

On the way towards the model of equation (1), we achieved a tunable $H_z(t)$; but the Josephson coupling is fixed, $H_x(t) = E_j/2$. Still, single-bit operations can be performed by controlling the bias voltage V_x (ref. 10). Furthermore, when the qubits are connected in parallel with an inductor (as in Fig. 3), the common LC-oscillator mode provides a two-bit coupling with weak, but constant $J^{ij} \approx (C^2/C_j^2)(E_j^2 L/\Phi_0^2)$, where $\Phi_0 = h/2e$. This coupling provides a two-bit gate if two qubits, i and j , are brought into resonance by biasing them with the same gate voltage $V_{xi} = V_{xj}$. Out of resonance, the two-bit coupling provides only a weak perturbation.

The external voltage source is part of a dissipative circuit with effective resistance R_V . Its Johnson–Nyquist voltage fluctuations destroy the phase coherence. The dephasing rate varies slightly during manipulations^{21,22}. At the degeneracy point, the decoherence time is:

$$\tau_V = \frac{1}{4\pi} \frac{R_K}{R_V} \left(\frac{C_j}{C_{qb}} \right)^2 \frac{\hbar}{E_j} \tanh \left(\frac{E_j}{2k_B T} \right) \quad (4)$$

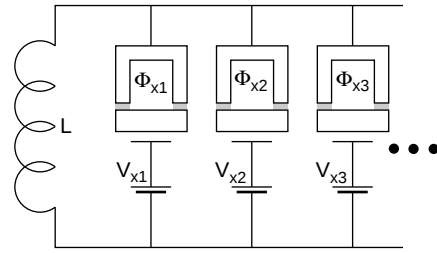


Figure 3 Design of a quantum computer. The coupling of the qubits is provided by the LC-oscillator mode in circuit shown. If the frequency of the LC-mode in the resulting circuit is large, $\hbar\omega_{LC} = \hbar(NC_{qb}L)^{-1/2} \gg E_j, E_{ch}, k_B T$, the fast oscillations produce an effective coupling of the qubits. We note that the system can be scaled to large numbers of qubits. In the idle state all effective Josephson couplings are tuned to zero, and the voltages are chosen such that the charge states are degenerate. Single-bit operations are performed by changing the gate voltage or flux of one qubit at a time. Two-bit operations between any two qubits are triggered by turning on the corresponding two Josephson couplings. The two lowest states of the qubit are separated from higher states, which exist in the real system, by the energies $E_C, \hbar\omega_{LC}, \hbar\omega_\Phi$. These should be larger than the energy scales of the qubit, $E_j, E_{ch}, k_B T$. If, in addition, switching processes of V_x and Φ_x are slow on the corresponding timescales, the requirements presented above also ensure that the higher states are not excited. Alternatively, instead of sudden switching, one can change the biases adiabatically. Another advantage of the presented design is that the result of a single-bit operation depends only on the time integral of energies $E_i(t)$ and $E_{ch}(t)$ over the operation period, but not on the profile of their time dependences.

Here R_V is compared to the quantum resistance $R_K = h/e^2 \approx 26 \text{ k}\Omega$. A small gate capacitance $C \approx C_{qb} \ll C_j$ helps further decoupling of the qubit from the environment. Both can be optimized to yield a phase coherence time that is long compared to typical operation times $\hbar/E_j$.

A problem with the simple design is that the eigenstates of the hamiltonian shown in equation (3) are non-degenerate at all V_x . Therefore, the relative phase of two logical states evolves even during idle periods. We can still store quantum information in the qubit, as becomes apparent after a transformation to the interaction representation. But this introduces an explicit time dependence in the operators, with the result that the unitary transformations not only depend on the time span τ of the operations but also on the time t_0 when they start. Hence the time elapsed since the beginning of the computation, multiplied by the energy spacing between the logical states should be controlled with high accuracy. A second problem of the simple design is the non-vanishing two-bit coupling, even out of resonance. It introduces an error in the computation. The design discussed below overcomes both these problems.

A crucial step towards the ideal model (equation (1)) is to tune the Josephson coupling. This is achieved in the design of Fig. 1b, where each Josephson junction is replaced by a d.c.-SQUID (see, for example, ref. 20). The SQUID is biased by an external flux Φ_x , coupled into the system through an inductor loop. If the loop self-inductance L_Φ is low the SQUID-controlled qubit is described by a hamiltonian of the form of equation (2), but with potential energy $2E_j \cos(\pi\Phi_x/\Phi_0) \cos \gamma$. Hence, the effective Josephson coupling is tunable by the external flux Φ_x between $2E_j^0$ and zero:

$$E_j(\Phi_x) = 2E_j^0 \cos(\pi\Phi_x/\Phi_0) \quad (5)$$

The SQUID-controlled qubit is described by the first two terms of the model hamiltonian shown in equation (1), with z - and x -components controlled independently by the gate voltage and the flux. In the idle state we keep $V_x = V_{deg}$ and $\Phi_x = \Phi_0/2$, so that the hamiltonian $\hat{H} = 0$. Changing one of them generates z - or x -rotations, respectively, that is, the elementary one-qubit operations.

With the improved design there is no need to control the total operation time t_0 , while the time dependence of the voltage and flux can be optimized such that the time span of the manipulations τ is long enough to simplify time control and short enough to speed up the computation.

Also, the circuit of the current source, with resistance R_b , which couples the flux Φ_x to the SQUID by the mutual inductance M , introduces fluctuations and may destroy the coherence of the qubit dynamics. At the degeneracy point, the decoherence time is^{21,22} $\tau_i = (1/\pi^3)(R_i/R_K)[\Phi_0^2/(E_J^0 M)]^2(\hbar/k_B T)$. This dephasing is slow if the current source is coupled weakly to the qubit (small M) and its resistance is high.

The control of the Josephson energies $E_J(\Phi_{xi})$ provides the possibility of coupling each selected pair of qubits, while keeping all the other ones uncoupled, bringing us close to the ideal model of equation (1). The simplest implementation of the coupling is to connect all N qubits in parallel with each other, and with inductor L (Fig. 3). Fast oscillations in the resulting LC-circuit produce an effective coupling of the qubits

$$\hat{H}_{\text{int}} = - \sum_{i < j} \frac{E_J(\Phi_{xi})E_J(\Phi_{xj})}{E_L} \hat{\sigma}_y^i \hat{\sigma}_y^j \quad (6)$$

where $E_L = [\Phi_0^2/(\pi^2 L)](C_J/C_{\text{qb}})^2$. The coupling shown in equation (6) can be understood as the magnetic energy of the inductor which is biased by a current composed of contributions from all qubits, $I^i \propto E_J^i \hat{\sigma}_y^i$.

With this design we can perform all gate operations. In the idle state the interaction hamiltonian of equation (6) is zero as all the Josephson couplings are turned off. The same is true during a one-qubit operation, as long as we perform one such operation at a time that is, only one $E_J^i \neq 0$. To perform a two-qubit operation with any given pair of qubits, say 1 and 2, E_J^1 and E_J^2 are switched on simultaneously, yielding the total hamiltonian $\hat{H} = -(E_J^1/2)\hat{\sigma}_x^1 - (E_J^2/2)\hat{\sigma}_x^2 - (E_J^1 E_J^2/E_L)\hat{\sigma}_y^1 \hat{\sigma}_y^2$. Although not identical to equation (1), these two-bit gates, in combination with the single-bit operations discussed above, also provide a complete set of gates required for quantum computation.

To demonstrate that the constraints on the set of system parameters can be met by available technology, we suggest a suitable set. We choose junctions with capacitance $C_J = 300$ aF, corresponding to a charging energy (in temperature units) $E_C \approx 3$ K, and a smaller gate capacitance $C = 30$ aF to reduce the coupling to the environment (even lower C are available and improve the performance further). The superconducting gap has to be slightly larger, $\Delta > E_C$. Thus at a working temperature of the order of $T = 50$ mK, the initial thermalization is assured. We further choose $E_J^0 = 50$ mK; so the timescale of one-qubit operations is $\tau_{\text{op}} = \hbar/E_J \approx 70$ ps. Fluctuations associated with the gate voltages (equation (4)), with resistance $R_V \approx 50 \Omega$, limit the coherence time to $\tau_V/\tau_{\text{op}} \approx 4,000$ operations. With the parameters of the flux-circuit $L_\Phi = 0.1$ nH, $M = 1$ nH and $R_f = 10^2 - 10^6 \Omega$, current fluctuations have a weak dephasing effect. To assure fast two-bit operations, we choose the energy scale E_L to be of the order of $10E_J$, which is achieved for $L \approx 3 \mu\text{H}$. With these parameters, the number of qubits in the circuit can be chosen in the range of 10–50, of course at the expense of shorter coherence times τ_V/N .

Some further remarks are in order.

(1) After the gate operations, the resulting quantum state has to be read out. This can be achieved by coupling a normal-state single-electron transistor capacitively to a qubit. The important aspect is that during computation the transistor is kept in a zero-current state and adds only to the total capacitance. When the transport voltage is turned on, the phase coherence of the qubit is destroyed, and the dissipative current in the transistor, which depends on the state of the qubit, can be read out. This quantum measurement process has been described explicitly in ref. 16 by an analysis of the time-evolution of the density matrix of the coupled system.

(2) Inaccuracy in the control of fluxes, voltages and the time-span of operations leads to diffusion of the actual quantum state from the one that exists in the absence of errors²³. A random error of order ϵ per gate limits the number of operations to a value which is of order ϵ^{-2} . For the circuit parameters above, $\epsilon = 1\%$ would lead to smaller effects than those produced by environment.

(3) Many powerful quantum algorithms make use of parallel operations on different qubits. Although this is not possible with the present system, it may be achievable by a more advanced design, making use of further tunable SQUIDs decoupling different parts of the circuit. Such modifications, as well as the further progress of nanotechnology, should provide longer coherence times and allow scaling to larger numbers of qubits. $\square$

Received 16 October 1998; accepted 12 January 1999.

- Lloyd, S. A potentially realizable quantum computer. *Science* **261**, 1569–1571 (1993).
- Bennett, C. H. Quantum information and computation. *Phys. Today* **48**(10), 24–30 (1995).
- DiVincenzo, D. P. Quantum computation. *Science* **269**, 255–261 (1995).
- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- King, B. E. *et al.* Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).
- Chuang, I. L., Gershenfeld, N. A. & Kubinec, M. Experimental implementation of fast quantum searching. *Phys. Rev. Lett.* **80**, 3408–3411 (1998).
- Cory, D. G. *et al.* Experimental error correction. *Phys. Rev. Lett.* **81**, 2152–2155 (1998).
- Jones, J. A., Mosca, M. & Hansen, R. H. Implementation of a quantum search algorithm on a quantum computer. *Nature* **393**, 344–346 (1998).
- Turchette, Q. A., Hood, C. J., Lange, W., Mabuchi, H. & Kimble, W. Measurement of conditional phase shifts for quantum logic. *Phys. Rev. Lett.* **75**, 4710–4713 (1995).
- Shnirman, A., Schön, G. & Hermon, Z. Quantum manipulations of small Josephson junctions. *Phys. Rev. Lett.* **79**, 2371–2374 (1997).
- Averin, D. V. Adiabatic quantum computation with Cooper pairs. *Solid State Commun.* **105**, 659–664 (1998).
- Bouchiat, V., Vion, D., Joyez, P., Esteve, D. & Devoret, M. H. Quantum coherence with a single Cooper pair. *Phys. Scripta* **T 76**, 165–170 (1998).
- Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Barenco, A. *et al.* Elementary gates for quantum computation. *Phys. Rev. A* **52**, 3457–3467 (1995).
- Shnirman, A. & Schön, G. Quantum measurements performed with a single-electron transistor. *Phys. Rev. B* **57**, 15400–15407 (1998).
- Rouse, R., Han, S. & Lukens, J. E. Observation of resonant tunneling between macroscopically distinct quantum levels. *Phys. Rev. Lett.* **75**, 1614–1617 (1995).
- Maassen v.d. Brink, A., Schön, G. & Geerligs, L. J. Combined single-electron and coherent-Cooper-pair tunneling in voltage-biased Josephson junctions. *Phys. Rev. Lett.* **67**, 3030–3033 (1991).
- Nakamura, Y., Chen, C. D. & Tsai, J. S. Spectroscopy of energy level splitting between two macroscopic quantum states of charge coherently superposed by Josephson coupling. *Phys. Rev. Lett.* **79**, 2328–2331 (1997).
- Tinkham, M. *Introduction to Superconductivity* (McGraw-Hill, New York, 1996).
- Leggett, A. J. *et al.* Dynamics of the dissipative two-state system. *Rev. Mod. Phys.* **59**, 1–85 (1987).
- Weiss, U. *Quantum Dissipative Systems* (World Scientific, Singapore, 1993).
- Miquel, C., Paz, J. P. & Zurek, W. H. Quantum computation with phase drift errors. *Phys. Rev. Lett.* **78**, 3971–3974 (1997).

Acknowledgements. We thank T. Beth, M. Devoret, D. P. DiVincenzo, E. Knill, K. K. Likharev and J. E. Mooij for discussions.

Correspondence and requests for materials should be addressed to Y.M. (e-mail: makhlin@tfp.physik.uni-karlsruhe.de).

Evidence against ‘ultrahard’ thermal turbulence at very high Rayleigh numbers

James A. Glazier*, Takehiko Segawa†, Antoine Naert‡ & Masaki Sano†

* Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556–5670, USA

† Research Institute of Electrical Communication, Tohoku University, 2-2-1 Katahira, Aoba-ku, Sendai 980, Japan

‡ Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 Allée d’Italie, 69364 Lyon Cedex 7, France

Several theories^{1–5} predict that a limiting and universal turbulent regime—‘ultrahard’ turbulence—should occur at large Rayleigh numbers (Ra, the ratio between thermal driving and viscous dissipative forces) in Rayleigh–Bénard thermal convection in a closed, rigid-walled cell. In this regime, viscosity becomes negligible, gravitationally driven buoyant plumes transport the heat

and the thermal boundary layer, where the temperature profile is linear, controls the rate of thermal transport. The ultrahard state is predicted to support more efficient thermal transport than 'hard' (fully developed) turbulence: transport efficiency in the ultrahard state grows as $Ra^{1/2}$, as opposed to $Ra^{2/7}$ in the hard state⁶. The detection of a transition to the ultrahard state has been claimed in recent experiments using mercury⁷ and gaseous helium⁸. Here we report experiments on Rayleigh–Bénard convection in mercury at high effective Rayleigh numbers, in which we see no evidence of a transition to an ultrahard state. Our results suggest that the limiting state of thermal turbulence at high Rayleigh numbers is ordinary hard turbulence.

In thermal turbulence, several dimensionless numbers determine the degree of turbulence. The most important of these is the Rayleigh number, Ra , the ratio between thermal driving and viscous dissipation forces: $Ra = \alpha g \Delta T l^3 / \kappa \nu$, where α is the coefficient of thermal expansion, κ the thermal conductivity, ν the kinematic viscosity, g the gravitational acceleration, l the typical length and ΔT the temperature difference across the cell. The second important quantity is the dimensionless velocity, the Reynolds number, Re , the ratio of shear forces to viscous forces: $Re = v l / \nu$, where v is a typical velocity. The fluid's Prandtl number, Pr , is the ratio between thermal and viscous dissipation: $Pr = \kappa / \nu$. Typical gases have $Pr \approx 1$, whereas $Pr > 1$ for most liquids (such as water and oil). The actual flow pattern of the fluid can be described in an averaged sense by the Nusselt number, Nu , the thermal transport efficiency, which is the ratio between gross thermal transport (including advection and diffusion) and diffusive thermal transport in the absence of flow. The larger Nu , the more efficient the convective thermal transport.

Convective thermal turbulence is inevitably anisotropic and non-homogeneous at large length scales. Two types of boundary layers are important. Near the container walls, flow velocity vanishes due to the no-slip condition which creates a highly sheared viscous boundary layer. The temperature profile becomes linear near the top and bottom walls, because only diffusion transports heat where advection is suppressed. The time-averaged temperature in the bulk is constant and equal to the average of the top and bottom plate temperatures due to strong mixing by the turbulent flow. The time-averaged profile is steep and linear in the thermal boundary layer. The thickness of the thermal boundary layer, which is the inverse of the temperature gradient, limits the gross heat transport across the turbulent cell. Thus Nu is proportional to the ratio between the cell height and the thickness of the thermal boundary layer.

The theories predicting $Nu \propto Ra^{1/2}$ make one of two assumptions. The first is that, because the thickness of the viscous boundary layer varies as $\sim Ra^{-1/2}$, it becomes negligible at very high Ra . Thus heat is advected by buoyant structures (for instance, plumes or thermals, rising masses of hot fluid or falling masses of cold fluid of classical mushroom-cloud shape) which move at the free fall velocity. That is, as viscous forces are negligible compared with inertial forces in this regime, the thermals accelerate as if they were free, undamped particles, subject to the buoyancy force produced by gravity. Because the free fall velocity, V scales as $\sim (\alpha g \Delta T) \propto Ra^{1/2}$, the heat flux scales as $Nu \propto Ra^{1/2}$. The other argument is that when the viscous boundary layer becomes thinner than the thermal boundary layer (about $Ra = 10^{14}$ for He⁹ and $Ra = 10^5$ for Hg^{7,10}), the thickness of the viscous boundary layer limits the heat flux, which also gives $Nu \propto Ra^{1/2}$. The critical Ra for the transition, Ra_{crit} , is predicted to depend on the Prandtl number of the fluid as $Ra_{crit} \propto Pr^4$ (ref. 1). We emphasize that whereas Ra_{crit} depends on Pr and the particular cell geometry, both Kraichnan, and Shraiman and Siggia, claim that the existence of the transition should be universal and independent of these factors^{1,3}. Extensive experiments by Wu have confirmed that scaling relations in turbulence are essentially independent of aspect ratio for aspect ratios above 1/2 (ref. 11). Shraiman and Siggia predict^{1,2} $Ra_{crit} \approx 10^6$ – 10^8 for Hg^{1,2}.

Cioni *et al.* proposed² modifications to the theory which predicted $Ra_{crit} = 5 \times 10^5$ in Hg⁷. The Libchaber experiments^{3,11–15} on turbulence in Rayleigh–Bénard convection in low temperature gaseous He found no indication of a shift in Nu scaling up to $Ra \approx 10^{14}$, but did see a suggestive change in the temperature power spectrum.

Because $Pr = 0.025$ at 20 °C in Hg compared to $Pr = 0.7$ for He gas below 1 atm at 5 K, searching for a new range in Hg required only a Ra number of the order of 10^8 . Our initial Hg experiments used cylindrical cells, with aspect ratios of one-half, one, and two, to reach $Ra \approx 2 \times 10^9$. Our results were compatible with the hard turbulence Nu -versus- Ra scaling exponent of $2/7 = 0.285$ (ref. 16). The break in the data for an aspect ratio of two at $Ra \approx 2 \times 10^5$ results from a pattern competition instability in the flow¹⁰. For small $Ra \approx 10^6$ – 10^8 , strong steady bulk mean flow in the cell of aspect ratio one reduces the measured scaling exponent to 0.25 (ref. 17). The exponent increases towards $2/7$ with increasing Ra .

These measurements appear to invalidate both of the mechanisms proposed for the onset of ultrahard turbulence. The thermal and viscous boundary layers cross below $Ra \approx 10^5$, so we should definitely have seen the transition if the boundary crossing theory were correct. The boundary layer thickness argument also seems to fail: in Hg, after the onset of hard turbulence and when the viscous boundary layer crosses the thermal (above $Ra \approx 10^6$), our experiments have shown that both boundary layers shrink together with a nearly constant ratio and a scaling exponent of -0.20 ± 0.02 , different from the theoretical value of $-1/2$ (ref. 17).

Cioni *et al.*⁷, using a larger Hg cell of aspect ratio one, measured up to $Ra \approx 5 \times 10^9$. They claimed that the Nu -versus- Ra scaling exponent increased above $2/7$ at a single point at the very top of their Ra range, but could not measure the new value. Chavanne *et al.*⁸ studied Rayleigh–Bénard convection with an aspect ratio of one-half in gaseous and liquid He. In gas near the gas–liquid critical point they saw an increasing Nu -versus- Ra scaling exponent above $Ra \approx 10^{11}$ and measured an approximate power law of

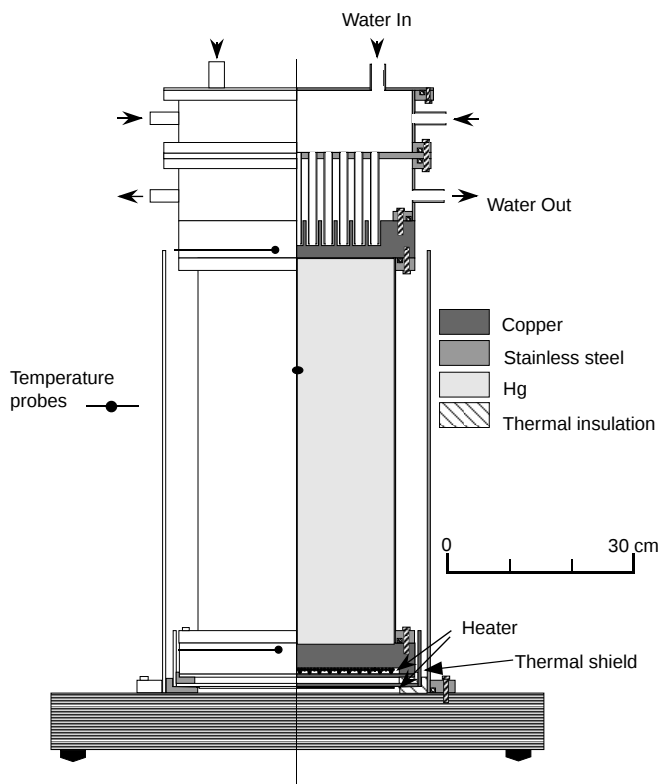


Figure 1 Schematic of the apparatus showing locations of bolometers.

$Nu \propto Pr^{0.072} Ra^{0.389 \pm 0.005}$. Chavanne *et al.* claimed that the new exponent was compatible with $1/2$ scaling due to a logarithmic correction.

Our current experiment uses a cylinder of aspect ratio one-half (30 cm by 60 cm). We show the apparatus in Fig. 1. The top and bottom plates are solid copper 5 cm thick, coupled on top to cooling water via 212 vertical, thermally anchored cooling pipes. Heating is supplied by four resistance coils soldered to the bottom plate and driven by four regulated direct current power supplies (GP110-30x4, Takasago, Kawasaki, Japan) with an accuracy of 0.01% and a maximum combined heat supply of $\sim 12,000$ W. Cooling water is supplied by two building-type air-conditioning units (capacity $\sim 20,000$ W) with proportional-integral-derivative (PID) temperature regulation controlled by a flow-rate throttle. Temperature control of the top plate is accurate to 2% of the total temperature difference (ΔT) at the maximum power. Top-plate temperature ranges from 15 °C to 42 °C and bottom-plate temperature from 18 °C to 130 °C. The maximum variation of Prandtl number occurs at maximum power when it is 0.024 at the top plate and 0.018 at the bottom plate. The horizontal temperature non-uniformity across the plate diameter is much less than 1%. The size and power capacity were dictated by our need to search three decades higher in Ra than the highest value predicted for the transition, so that we could definitively determine whether the transition exists.

A thermal shield surrounds the cell. It can be temperature-matched to the cell to eliminate radiative and convective heat losses. Without the thermal shield, heat loss from the bottom plate is less than 1% at the highest Ra and much lower for smaller Ra. With the thermal shield, the worst case heat loss is less than 0.1%. Thermal conduction through the cell walls is negligible. The power supplied by the heater thus gives the heat flux to an accuracy of better than 1%. The temperature probes are semiconductor bolometers individually calibrated to an accuracy of better than 10^{-3} °C. Because the top- and bottom-plate temperature probes are embedded in the plates 1 cm from the copper–Hg interface, we must correct for the temperature drop across the copper. Temperature drops between the thermistor embedded in the bottom plate and the top surface of that plate, and between the bottom surface of the top plate and the thermistor embedded in this plate, are calculated from the applied heat flux and the thermal conductivity of the copper ($401 \text{ J m}^{-1} \text{ K}^{-1}$), and are used to correct the value of ΔT measured by the top and bottom thermistors. Ra and Nu are calculated using the corrected ΔT (ref. 17).

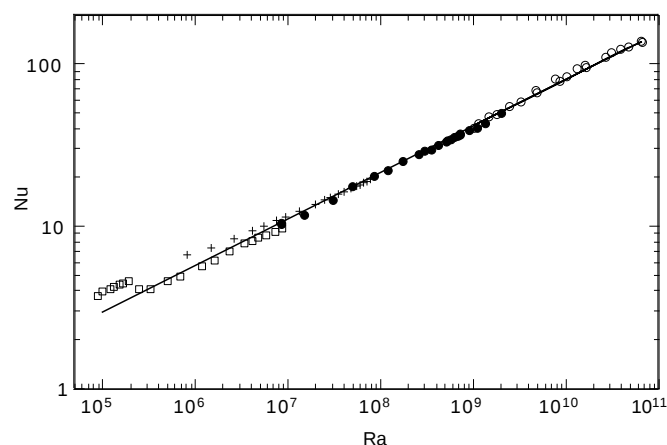


Figure 2 Nusselt number as a function of Rayleigh number. Open circles; aspect ratio = 0.5, large cell. Crosses; aspect ratio = 1, small cell. Squares; aspect ratio = 2, small cell. Filled circles; aspect ratio = 0.5, small cell. The solid line is $Nu \propto Ra^{0.285}$. The discontinuity in the data for aspect ratio = 2 is a result of a pattern competition instability. The reduced slope for small Ra for the data for an aspect ratio of one results from strong bulk circulation.

In Fig. 2 we show the Nu-versus-Ra curve for the combined data from our four experiments. From $Ra = 2 \times 10^5$ to 8×10^{10} , the Nusselt numbers lie on top of each other with a constant scaling exponent of 0.29 ± 0.01 . This Rayleigh number is, to our knowledge, the highest yet achieved in thermal convection in a fluid of low Prandtl number, and the estimated Reynolds number is 5×10^5 (ref. 17), which is higher than that of Chavanne *et al.* Even at the highest Ra, our data show no indication of an increase in power law.

The Nu-versus-Ra curve is a highly averaged characterization of the fluid flow. In particular, transitions which change the temperature histograms and power spectra (which have never been observed) might not change the Nu-versus-Ra exponent. Figure 3 compares the power spectra and histograms for temperature time series taken for $Ra = 1.85 \times 10^9$ and $Ra = 5.14 \times 10^{10}$. The best fit to the scaling range of the temperature power-spectrum gives a slope of -1.47 ± 0.07 . This slope and the range (less than five decades) is compatible with the result of Wu *et al.*¹⁴ obtained for convective hard turbulence in helium. Except for the expected increase in inertial scaling range and the shrinking width of exponential decay in the temperature histogram, the two curves correspond exactly to each other. The small asymmetry in the histogram shape results from the location of the temperature probe above the centre of the cell, 15 cm from the cold top plate¹⁰. Again we see no sign of a qualitative change in the turbulence.

Chavanne *et al.*⁸ clearly see an interesting effect in gaseous He—an increase of Nu at very high Ra—but the transition is gradual and never shows a clear power law, certainly not $Ra^{1/2}$. One complication in interpreting their results is that Pr changes rapidly near the critical point so that it is not constant from the top to the bottom of their cell. Although it is tempting to attribute this regime to ultrahard turbulence, it may instead be due to Pr effects, which partially mimic the predicted behaviour of ultrahard turbulence.

For our largest Ra, the Pr at the top plate (at $T = 130$ °C) is 0.018 and at the bottom plate $Pr = 0.024$ at $T = 42$ °C, a change of 35%. However the maximum Pr is 0.027, so viscous damping is always

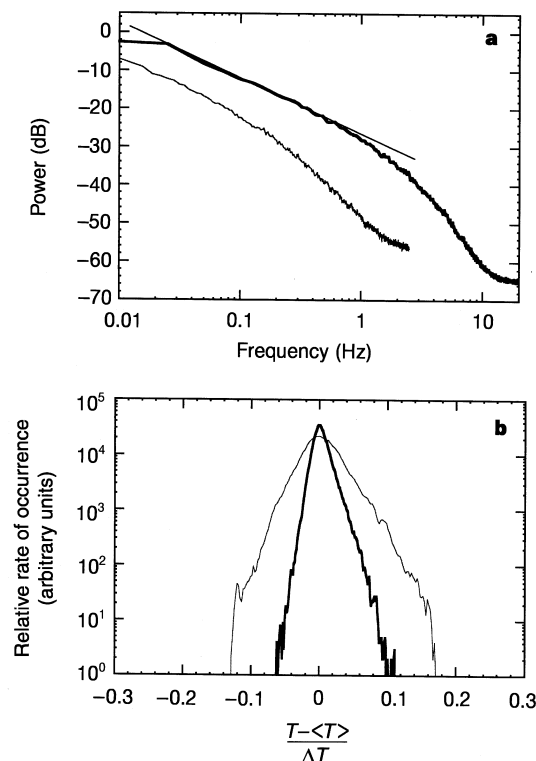


Figure 3 Temperature fluctuations data. **a**, Power spectra; **b**, histograms (thin line, $Ra = 1.85 \times 10^9$; thick line, $Ra = 5.14 \times 10^{10}$).

negligible. Thus the fundamental dissipation process remains the same and we expect the turbulence to be qualitatively unaffected by the Pr shift. In the He experiments the Pr shift is from 0.7 to 7 with the result that the primary dissipation mechanism shifts from thermal diffusion to viscous damping, for different locations within the same cell at a single Ra number.

In Hg, in which these complicating factors do not contribute (and for substantially higher Re), we see no evidence of a transition. Thus ultrahard turbulence may not exist. □

Received 17 July; accepted 29 December 1998.

- Shraiman, B. I. & Siggia, E. D. Heat transport in high-Rayleigh-number convection. *Phys. Rev. A* **42**, 3650–3653 (1990).
- Siggia, E. D. High Rayleigh number convection. *Annu. Rev. Fluid Mech.* **26**, 137–168 (1994).
- Kraichnan, R. H. Turbulent thermal convection at arbitrary Prandtl number. *Phys. Fluids* **5**, 1374–1389 (1962).
- Howard, L. N. Bounds on flow quantities. *Annu. Rev. Fluid Mech.* **4**, 473–494 (1972).
- Busse, F. H. The optimum theory of turbulence. *Adv. Appl. Mech.* **18**, 77–121 (1978).
- Castaing, B. *et al.* Scaling of hard thermal turbulence in Rayleigh–Bénard convection. *J. Fluid Mech.* **204**, 1–30 (1989).
- Cioni, S., Ciliberto, S. & Sommeria, J. Strongly turbulent Rayleigh–Bénard convection in mercury: comparison with results at moderate Prandtl number. *J. Fluid Mech.* **335**, 111–140 (1997).
- Chavanne, X. *et al.* Observation of the ultimate regime in Rayleigh–Bénard convection. *Phys. Rev. Lett.* **79**, 3648–3651 (1997).
- Belmonte, A., Tilgner, A. & Libchaber, A. Temperature and velocity boundary layers in turbulent convection. *Phys. Rev. E* **50**, 269–279 (1994).
- Segawa, T., Naert, A. & Sano, M. Matched boundary layers in turbulent Rayleigh–Bénard convection of mercury. *Phys. Rev. E* **57**, 557–560 (1998).
- Wu, X. Z. & Libchaber, A. Scaling relations in thermal turbulence: The aspect-ratio dependence. *Phys. Rev. A* **45**, 842–845 (1992).
- Schewe, P. F. & Stein, B. One of the most convective fluids observed in the laboratory. *Physics News Update* 342-1 (Am. Inst. Phys., 1997).
- Heslot, F., Castaing, B. & Libchaber, A. Transition to turbulence in helium gas. *Phys. Rev. A* **36**, 5870–5873 (1987).
- Wu, X. Z., Kadanoff, L., Libchaber, A. & Sano, M. Frequency power spectrum of temperature fluctuations in free convection. *Phys. Rev. Lett.* **64**, 2140–2143 (1990).
- Wu, X. Z. & Libchaber, A. Non-Boussinesq effects in free thermal convection. *Phys. Rev. A* **43**, 2833–2839 (1991).
- Takeshita, T., Segawa, T., Glazier, J. A. & Sano, M. Thermal turbulence in mercury. *Phys. Rev. Lett.* **76**, 1465–1468 (1995).
- Naert, A., Segawa, T. & Sano, M. High-Reynolds-number thermal turbulence in mercury. *Phys. Rev. E* **56**, 1302–1305 (1997).

Acknowledgements. J.A.G. was supported by the NSF and the ACS/PRF; M.S. acknowledges support from the Japanese Grant-in-Aid for Science Fund for the Ministry of Education, Science and Culture. Both are supported by the NSF-JSPS Cooperative Science Program.

Correspondence and requests for materials should be addressed to J.A.G. (e-mail: jglazier@rameau.phys.nd.edu).

Picosecond–milliångström lattice dynamics measured by ultrafast X-ray diffraction

Christoph Rose-Petruck*†, Ralph Jimenez*†, Ting Guo*, Andrea Cavalleri*, Craig W. Siders*, Ferenc Rákosi*†, Jeff A. Squier§, Barry C. Walker*†, Kent R. Wilson* & Christopher P. J. Barty‡

* Department of Chemistry and Biochemistry, ‡ The Institute for Nonlinear Science, § Department of Electrical and Computer Engineering, The University of California, San Diego, La Jolla, California 92093–0339, USA

Fundamental processes on the molecular level, such as vibrations and rotations in single molecules, liquids or crystal lattices and the breaking and formation of chemical bonds, occur on time-scales of femtoseconds to picoseconds. The electronic changes associated with such processes can be monitored in a time-resolved manner by ultrafast optical spectroscopic techniques¹, but the accompanying structural rearrangements have proved

more difficult to observe. Time-resolved X-ray diffraction has the potential to probe fast, atomic-scale motions^{2–5}. This is made possible by the generation of ultrashort X-ray pulses^{6–10}, and several X-ray studies of fast dynamics have been reported^{6–8,11–15}. Here we report the direct observation of coherent acoustic phonon propagation in crystalline gallium arsenide using a non-thermal, ultrafast-laser-driven plasma—a high-brightness, laboratory-scale source of subpicosecond X-ray pulses^{16–19}. We are able to follow a 100-ps coherent acoustic pulse, generated through optical excitation of the crystal surface, as it propagates through the X-ray penetration depth. The time-resolved diffraction data are in excellent agreement with theoretical predictions for coherent phonon excitation²⁰ in solids, demonstrating that it is possible to obtain quantitative information on atomic motions in bulk media during picosecond-scale lattice dynamics.

Crystalline gallium arsenide (GaAs) is available in large samples of very high crystalline quality, making it an ideal system for quantitative investigations using optical pumping and X-ray probing. Its physical parameters are known with great precision²¹, and numerous prior studies²² on its ultrafast electronic properties provide a solid foundation for interpretation, and for testing the potential of ultrafast X-ray diffraction. Indeed, some information on ultrafast lattice dynamics after optical excitation has already been indirectly inferred from a variety of linear and nonlinear optical techniques^{22–24}. But owing to the short penetration depth of visible light, information on bulk dynamics in absorbing materials has not yet been obtained.

Figure 1 shows a schematic of our experiment. An ultrashort pulse of laser-generated Cu K α X-rays (consisting of two closely spaced lines, K α_1 and K α_2) diffracts in a symmetric Bragg configuration from the 3.26-Å (111) lattice spacing in GaAs, penetrating ~ 2 μ m into the bulk along the surface normal. A 30-fs pump pulse with variable time delay generates electron–hole pairs via interband excitation within the submicrometre penetration depth of the 800-nm light. By illuminating with light only a portion of the area probed by X-rays, we simultaneously observe the K α lines from both photopumped and unperturbed areas of the semiconductor surface. Two-minute exposure X-ray-CCD images are normalized with respect to the incident X-ray flux and binned within the region of uniform illumination. In Fig. 2, the measured and calculated (described below) diffraction profiles are shown as a function of angular deviation from the Bragg angle, $\theta - \theta_B$, and of pump–probe time delay. Zero delay corresponds to the initial deviation of the diffracted signal. We observe for these early times that both of the original K α lines broaden and shift slightly to larger angles (~ 20 arcsec), while two new lines appear, broader and weaker than the original lines and deviate by approximately ~ 150 arcsec relative to the original Bragg angle. As the pump–probe delay is increased, these new lines decrease in width, increase in intensity, and merge asymptotically with the still broadened and shifted main lines. Finally, the angle-integrated diffraction signal (not shown) increases monotonically with delay, reaching a plateau of twice the unperturbed value.

The incident optical energy couples into the material by promoting electrons from the valence to the conduction band. Single- and two-photon (as well as free-carrier) absorption contribute to excitation during absorption of the pump pulse. Bandgap renormalization and state filling, not well characterized at these fluences, introduce additional nonlinearities to the excitation process. Also, efficient carrier diffusion at early times may smooth the energy deposition profile in a few picoseconds. After absorption, energy is transferred to the lattice via intraband relaxation²² and delayed Auger heating²⁵. Existing estimates of the Auger coefficient²⁶ in GaAs indicate that most of the energy is efficiently transferred to the lattice within a few picoseconds. X-ray diffraction, primarily sensitive to the acoustic phonon population, occurs only after the decay of the initially generated longitudinal optical phonons²². Following

† Present addresses: Department of Chemistry, Brown University, Providence, Rhode Island, USA (C.R.-P.); The Scripps Research Institute, La Jolla, California, USA (R.J.); IMRA America, Ann Arbor, Michigan, USA (F.R.); Department of Physics and Astronomy, the University of Delaware, Newark, Delaware, USA (B.C.W.).

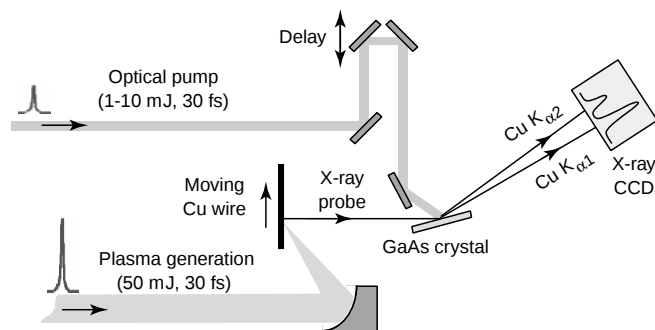


Figure 1 Diagram of the table-top ultrafast X-ray diffractometer. The 800-nm laser pulses are generated at 20 Hz by an ultrafast Ti:sapphire laser producing 30-fs pulses of 200-mJ energy³⁰. The output of the laser is used for both sample excitation and X-ray generation. The latter is done by focusing 30-fs, 50-mJ, laser pulses onto a moving Cu wire in vacuum, resulting in a source with measured diameter of $\sim 25\text{ }\mu\text{m}$. With an average laser power of 1.0 W, $\sim 5 \times 10^{10}$ Cu K α photons ($4\pi\text{ sr s}^{-1}$) are produced. The emitted K α radiation—which consists of two closely spaced, K α_1 (1.5407 Å) and K α_2 (1.5439 Å), lines—is diffracted from the optically pumped GaAs and detected by an X-ray CCD camera, with a data acquisition time of two minutes per delay step. We excite the GaAs crystal with ~ 30 -fs pulses at a fluence of 59 mJ cm^{-2} . The flat-top pump beam size is set to $\sim 4\text{ mm}^2$ ($95 \pm 5\%$ of peak) so that it includes the sample area encompassing the Bragg angles of the two K α lines and the auxiliary lines. The crystal is continuously moved during the experiment in order to minimize the effects of cumulative damage by the optical pump pulse.

our line of interpretation and the available parameters, we calculate a ~ 300 -nm thermal energy deposition length, with a maximum surface temperature increase of $\sim 800\text{ K}$ after 10 ps. Therefore, the efficient transfer of incident optical energy to lattice thermal energy can be considered complete after ~ 10 ps.

Subsequent lattice dynamics and related propagation effects are interpreted as proposed by Thomsen *et al.*²⁰, who solved the elastic equations assuming impulsively generated thermal stress in an absorbing solid. Stress is relieved by lattice expansion that necessa-

rily starts at the crystal surface, with Newton's third law subsequently driving a travelling compression/expansion strain wave into the bulk at the longitudinal speed of sound ($v_L = 5,397\text{ m s}^{-1}$ in (111)-GaAs²¹). Hence, mechanical relaxation of an impulsively stressed lattice layer of thickness d occurs within a time span of dv_L^{-1} , which corresponds to ~ 300 ps for our X-ray probe depth. As both energy deposition and probing occur significantly faster than mechanical relaxation, this is a regime where the generation and observation of coherent lattice dynamics is possible. This contrasts with many prior experiments^{6,13}, where laser heating is slower than mechanical relaxation and the strain is linearly proportional to the temperature. We note that for our experiment, a fully thermo-elastic treatment with coherent phonon excitation must be applied.

We numerically solve the thermo-elastic problem using the known parameters²¹ for GaAs, together with the calculated energy deposition depth and temperature rise (Fig. 3). Only three parameters are used in the model: the peak strain which is obtained directly from the data, the acoustic velocity which is known from the literature²¹, and the absorption length, ζ , which we estimate to be 300 nm. We systematically varied ζ within its uncertainty ($\pm 30\%$); 275 nm gave the best fit to the data. The calculated temperature increase corresponds to a maximum surface strain of $\sim 0.25\%$, or 8 mÅ of lattice expansion, after the stress is released, which is in good agreement with the experimental 150-arcsec deviation seen near zero delay in Fig. 2a. However, our data cannot exclude additional non-thermal strain generation²⁰ from the hot, dense electron-hole plasma at early times. A 100-ps acoustic pulse propagates into the crystal at v_L and has peak bipolar strain of approximately $\pm 0.12\%$, or $\pm 3\text{ mÅ}$. The theoretical time-resolved diffraction curves of Fig. 2b, corresponding to the strain profile in Fig. 3, were calculated using standard Takagi-Taupin^{27–29} techniques and convolved with the spectral and spatial distributions of our X-ray source. They are in good agreement with the experimental data of Fig. 2a. Inclusion of thermal diffusion²⁰ into the elastic equations effected higher-order dispersive strain behaviour, but, to within our experimental resolution, produces insignificant changes to the calculated diffraction profiles.

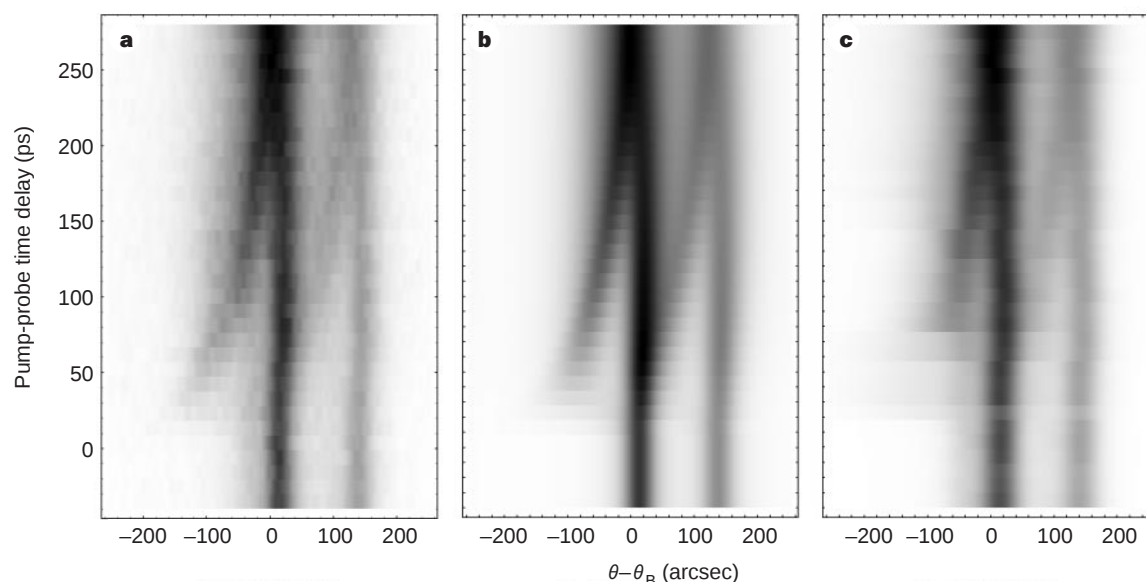


Figure 2 Experimentally measured (a), theoretically calculated (b), and iteratively inverted (c) time- and angle-resolved diffraction curves for optically excited GaAs. The horizontal axis is angular deviation from the K α_1 Bragg angle calculated assuming unit refractive index. Vertical axis is pump-probe time delay, with the zero of delay being chosen to coincide with the initial deviation of the lines from their unpumped appearance. For c, the iterative genetic-algorithm inversion

procedure varied the depth-dependent strain, defined by a 200-nm-spaced, cubic-spline-interpolated grid, and calculated the corresponding diffraction patterns until a good fit was achieved. Five inversions with different parameters of the genetic algorithm, such as random number seed, population size, crossover probability, and mutation rate, were performed and subsequently averaged.

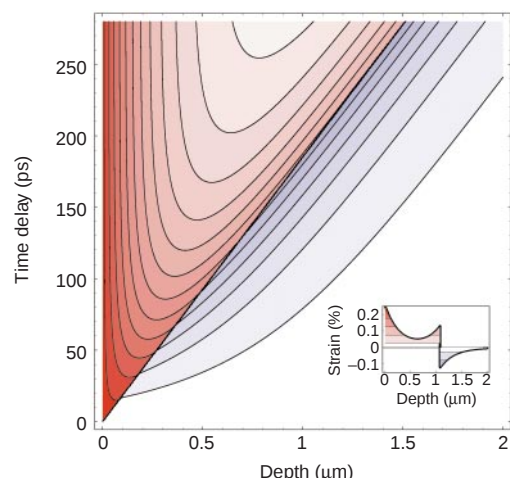


Figure 3 Theoretically calculated percentage lattice expansion, or strain, within the GaAs crystal as a function of depth and pump-probe time delay. Inset, the depth-dependent strain at a pump-probe delay of 200 ps. As indicated in the inset, regions of lattice expansion are indicated with an increasingly strong red, while regions of lattice compression are indicated similarly in blue. The white region indicates zero strain, that is, the unpumped lattice spacing. Contours are drawn at intervals of 0.02% of the static lattice plane spacing ($d_0 = 3.26 \text{ \AA}$), with the peak strain at the surface corresponding to 0.25%, that is $8.2 \text{ m\AA}$ of lattice expansion. The prominent red and blue diagonal represents the elastic strain wave, which starts at the surface and propagates, with compression leading expansion, into the bulk at $5,397 \text{ m s}^{-1}$.

Comparing Figs 2a, b and 3, we see that the initially thin but highly strained layer of lattice expansion produces the broad, low-intensity lines at early time delays. As the strain wave propagates away from the surface, the layer of surface expansion thickens and these lines correspondingly become narrower and more intense, but diffract at smaller angular deviations due to the weaker average strain. In parallel, the compression wave leading the propagation into the bulk contributes to the slight shift to higher angles seen in the main $K\alpha$ lines. At the longest delays, the strain wave has largely moved beyond the depths probed by the X-ray pulse, and we see diffraction lines broadened and shifted slightly to lower angles due to the exponential surface strain of the relaxed lattice. The strain for these late times is simply proportional to the temperature distribution. Finally, the theoretical angle-integrated diffraction signal similarly reproduces the monotonic increase and plateau behaviour seen in the experimental data.

Additionally, we performed an iterative genetic-algorithm inversion, obtaining the strain from the measured data. The angle- and time-resolved diffraction curves corresponding to the retrieved strain are shown in Fig. 2c. Spectral and geometrical broadening of the diffraction lines significantly limits the uniqueness of the result. Nevertheless, we obtain qualitatively similar strain behaviour to that shown in Fig. 3. We retrieve an exponential surface strain, with $8\text{-m\AA}$ peak strain, and a unipolar, $3\text{-m\AA}$ strain pulse which propagates into the bulk at several thousand metres per second. Although the physical model that we used is fully consistent with the measured data (to within our experimental resolution), the spectral and geometrical broadening mentioned above mean that it is not a unique interpretation of that data.

This work, using a table-top laboratory apparatus, demonstrates with a simple crystalline material the direct observation of milliangström atomic motion on the picosecond timescale. Further developments of the technique should permit direct observation by ultrafast X-ray diffraction of the ultrafast atomic motions accompanying a wide variety of physical, chemical, and perhaps biological processes. □

Received 21 October 1998; accepted 9 February 1999.

1. Elsaesser, T., Fujimoto, J. G., Wiersma, D. A. & Zinth, W. (eds) *Ultrafast Phenomena XI* (Springer, Berlin, 1998).
2. Rentzepis, P. M. & Helliwell, J. R. (eds) *Time-Resolved Diffraction* (Oxford Univ. Press, 1997).
3. Bergsma, J. P. *et al.* Transient x-ray scattering calculated from molecular dynamics. *J. Chem. Phys.* **84**, 6151–6160 (1986).
4. Wark, J. Time-resolved x-ray diffraction. *Contemp. Phys.* **37**, 205–218 (1996).
5. Ben-Nun, M., Cao, J. S. & Wilson, K. R. Ultrafast x-ray and electron diffraction: Theoretical considerations. *J. Phys. Chem. A* **101**, 8743–8761 (1997).
6. Larson, B. C., Tischler, J. Z. & Mills, D. M. Nanosecond resolution time-resolved x-ray study of silicon during pulsed-laser irradiation. *J. Mater. Res.* **1**, 144–154 (1986).
7. Srajer, V. *et al.* Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science* **274**, 1726–1729 (1996).
8. Chen, P., Tomov, I. V. & Rentzepis, P. M. Time resolved heat propagation in gold crystals by means of picosecond x-ray diffraction. *J. Chem. Phys.* **104**, 10001–10007 (1996).
9. Schoenlein, R. W. *et al.* Femtosecond x-ray pulses at $0.4 \text{ \AA}$ generated by 90 degree Thomson scattering: a tool for probing the structural dynamics of materials. *Science* **274**, 236–238 (1996).
10. Murnane, M. M., Kapteyn, H. C., Rosen, M. D. & Falcone, R. W. Ultrafast x-ray pulses from laser-produced pulses. *Science* **251**, 531–536 (1991).
11. Rischel, C. *et al.* Femtosecond time-resolved X-ray diffraction from laser-heated organic films. *Nature* **390**, 490–492 (1997).
12. Chin, A. H. *et al.* in *Ultrafast Phenomena XI* (eds Elsaesser, T., Fujimoto, J. G., Wiersma, D. A. & Zinth, W.) 401–403 (Springer, Berlin, 1998).
13. Buschert, J. R., Tischler, J. Z., Mills, D. M., Zhao, Q. & Corella, R. Time-resolved x-ray diffraction study of laser annealing in silicon at grazing-incidence. *J. Appl. Phys.* **66**, 3523–3525 (1989).
14. Wark, J. S., Riley, D., Woolsey, N. C., Keihn, G. & Whitlock, R. R. Direct measurement of compressive and tensile strain during shock breakout by use of subnanosecond x-ray diffraction. *J. Appl. Phys.* **68**, 4531–4534 (1990).
15. Larsson, J. *et al.* Ultrafast structural changes measured by time-resolved x-ray diffraction. *Appl. Phys. A* **66**, 587–591 (1998).
16. Kieffer, J.-C. *et al.* Ultrafast x-ray sources. *Phys. Fluids B* **5**, 2676–2681 (1993).
17. Rousse, A. *et al.* Efficient K alpha x-ray source from femtosecond laser-produced plasmas. *Phys. Rev. E* **50**, 2200–2207 (1994).
18. Kmetec, J. D. *et al.* MeV x-ray generation with a femtosecond laser. *Phys. Rev. Lett.* **68**, 1527–1530 (1992).
19. Gallant, P. *et al.* Subpicosecond time-resolved x-ray spectroscopy of plasmas produced by high intensity ultrashort laser pulses. *Proc. SPIE* **3157**, 44–51 (1997).
20. Thomsen, C., Grah, H. T., Maris, H. J. & Tauc, J. Surface generation and detection of phonons by picosecond light pulses. *Phys. Rev. B* **34**, 4129–4138 (1986).
21. Blakemore, J. S. Semiconducting and other major properties of gallium arsenide. *J. Appl. Phys.* **53**, R123–R181 (1981).
22. Shah, J. *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* (Springer, Berlin, 1996).
23. Shank, C. V., Yen, R. & Hirlmann, C. Femtosecond-time-resolved surface structural dynamics of optically excited silicon. *Phys. Rev. Lett.* **51**, 900–902 (1983).
24. Saeta, P., Wang, J.-K., Siegal, Y., Bloembergen, N. & Mazur, E. Ultrafast electronic disordering during the femtosecond laser melting of GaAs. *Phys. Rev. Lett.* **67**, 1023–1026 (1991).
25. Downer, M. C. & Shank, C. V. Ultrafast heating of silicon on sapphire by femtosecond optical pulses. *Phys. Rev. Lett.* **56**, 761–764 (1986).
26. Strauss, U. & Ruhle, W. Auger recombination in intrinsic GaAs. *Appl. Phys. Lett.* **62**, 55–57 (1993).
27. Takagi, S. A dynamical theory of diffraction for a distorted crystal. *J. Phys. Soc. Jpn* **26**, 1239–1253 (1969).
28. Taupin, D. Théorie dynamique de la diffraction des rayons X par les cristaux déformés. *Bull. Soc. Fr. Minér. Crist.* **87**, 469–511 (1964).
29. Larson, B. C., White, C. W., Noggle, T. S., Barhost, J. F. & Mills, D. M. Time-resolved x-ray diffraction measurement of the temperature and temperature gradients in silicon during pulsed laser annealing. *Appl. Phys. Lett.* **42**, 282–284 (1983).
30. Barty, C. P. J. *et al.* Generation of 18-fs, multiterawatt pulses using regenerative pulse shaping and chirped pulse amplification. *Opt. Lett.* **21**, 668–670 (1996).

Correspondence and requests for materials should be addressed to K.R.W. (e-mail: krwilson@ucsd.edu).

Directed nucleation of calcite at a crystal-imprinted polymer surface

S. M. D'Souza*, C. Alexander*, S. W. Carr†‡, A. M. Waller§, M. J. Whitcombe* & E. N. Vulfson*

* Institute of Food Research, Earley Gate, Whiteknights Road, Reading RG6 6BZ, UK

† Unilever Research, Port Sunlight Laboratory, Wirral L63 3JW, UK

§ Unilever Research, Weena 455, 3000DK Rotterdam, Netherlands

The finely tuned properties of natural biominerals and composites reflect the remarkable level of control that is exercised over the size, shape and organization of the constituent crystals^{1–4}. Achieving this degree of control over synthetic materials might therefore lead to superior material properties. Organic small molecules, polymers or surfactant mesophases have been used

‡ Present address: ANSTO, Private Mail Bag 1, Mendi 2234, Australia.

to guide the growth and morphology of inorganic materials via steric constraints or structure-directing interactions^{5–16}. Here we show that synthetic polymers can be imprinted with motifs of crystal surfaces so as to template the growth of specific crystal phases. Our polymers, imprinted with calcite, are able to induce the nucleation of calcite under conditions favouring the growth of aragonite (another polymorph of calcium carbonate). The synthesis of the polymers, based on the principles of molecular imprinting^{17–20}, involves the adsorption of functional monomers to a calcite surface, followed by co-polymerization with a cross-linker to create an imprint of the crystal surface. Subsequent removal of the calcite template yields a polymer matrix with a surface functionality mirroring the crystal face and able to promote the nucleation of calcite. We expect that the molecular-imprinting approach to directed nucleation can be applied to crystals other than calcite.

A schematic representation of the imprinting and directed nucleation processes reported here is shown in Fig. 1. We use 6-methacrylamidohexanoic acid (**1** in Fig. 1) as the functional monomer, and divinylbenzene (DVB) as the crosslinker, to convert calcite-templated monomer assemblies into a polymer matrix. In Fig. 2 we show scanning electron micrographs of the rhombohedral calcite crystals used as template (Fig. 2a), the polymer imprinted with these crystals (PI-1; Fig. 2b), and the polymer after dissolution of the crystals with mineral acids (Fig. 2c, d). The crystal-shaped ‘holes’ in the polymer matrix left after removal of the template are clearly visible in Fig. 2c and d.

Experiments to establish the ability of the imprinted polymer

matrix to promote the crystallization of calcite were performed in supersaturated solutions of calcium carbonate under two sets of conditions. Initially, solutions of salts (1.0 mM CaCl_2 , 0.8 mM Na_2CO_3) in pure water were used, employing 1.0 mg ml^{-1} of polymer. Under these conditions the formation of rhombohedral calcite at the imprint sites was observed, with most of the crystals being noticeably larger than those formed in solution in the absence of polymer (Fig. 2e, f). This was to be expected if nucleation in at least some imprinted sites occurred much faster than in solution, giving the crystals formed on the polymer more time to grow. Very few crystals were detected on the surface of control polymers CP-1 (prepared in the absence of crystals) and CP-2 (prepared with crystals but omitting monomer **1**) (Fig. 2g, h). The nucleation we observe at the imprinted polymer sites is unlikely to be the result of seeding by small undissolved crystals since CP-2, which does not nucleate, should retain the same number of seeds, having received identical treatment to PI-1. The polymers were also examined by high-resolution scanning electron microscopy, infrared microscopy, X-ray powder diffraction and Fourier-transform infrared spectroscopy (FT-IR). No polymer-bound calcite was detected by any of these techniques (see Fig. 3 for powder X-ray diffraction and FT-IR results), thus providing further evidence that the nucleation observed is indeed due to the imprinting effect rather than to residual calcite seed crystals somehow ‘surviving’ the acid wash. Similarly, energy dispersive X-ray scattering was able to detect calcium on imprinted polymers after calcite nucleation, but not before, confirming that residual calcite crystals are not responsible for the nucleation effects observed.

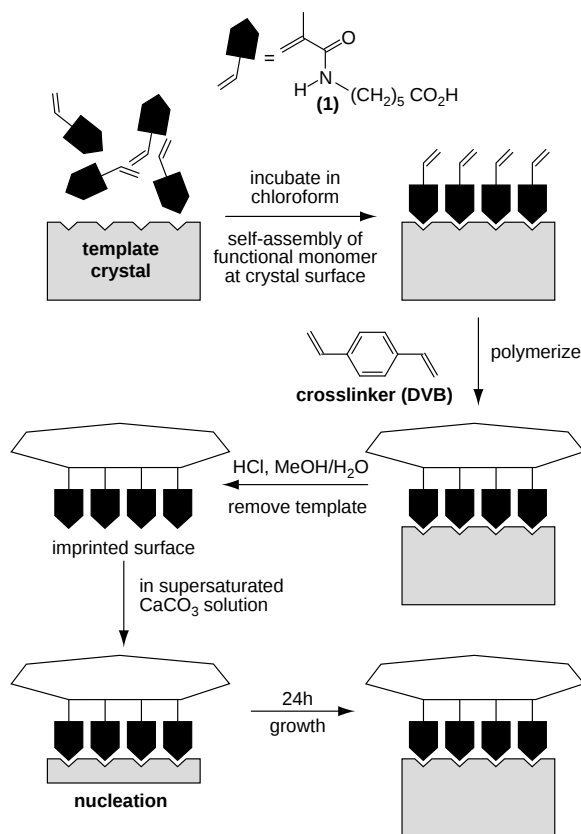


Figure 1 Schematic representation of the imprinting process and directed nucleation. Polymers were synthesized as follows. Calcite or aragonite (1.0 g) was stirred with 3 ml of a 26.7 mM solution of **1** in CHCl_3 in a sealed tube for 24 h. 1.29 g DVB (80% technical grade) was added with 39 mg azo-bis-cyclohexanecarbonitrile (1 mol% with respect to total polymerizable double bonds). The samples were degassed, sealed under reduced pressure and polymerized photochemically at 4°C. After gelation, the polymers were heated (65°C, 15 h) to complete the polymerization, dried *in vacuo*, and ground to fragments of size

~100 $\mu\text{m} \times 100 \mu\text{m}$. Template crystals were removed by washing with acidified aqueous methanol (1 M HCl, 5 $\times$ 100 ml), and the polymers were finally washed with acetonitrile and water and dried *in vacuo*. Control polymers were prepared via the same method: CP-1 contained **1**, DVB and CHCl_3 only, and CP-2 was prepared with calcite, CHCl_3 and DVB. (Template calcite and aragonite crystals were prepared by standard methods^{21,22}. 6-Methacrylamidohexanoic acid²³ (**1**) was prepared in 74% yield by the Schotten-Baumann procedure from 6-amino-hexanoic acid and methacrylic anhydride.)

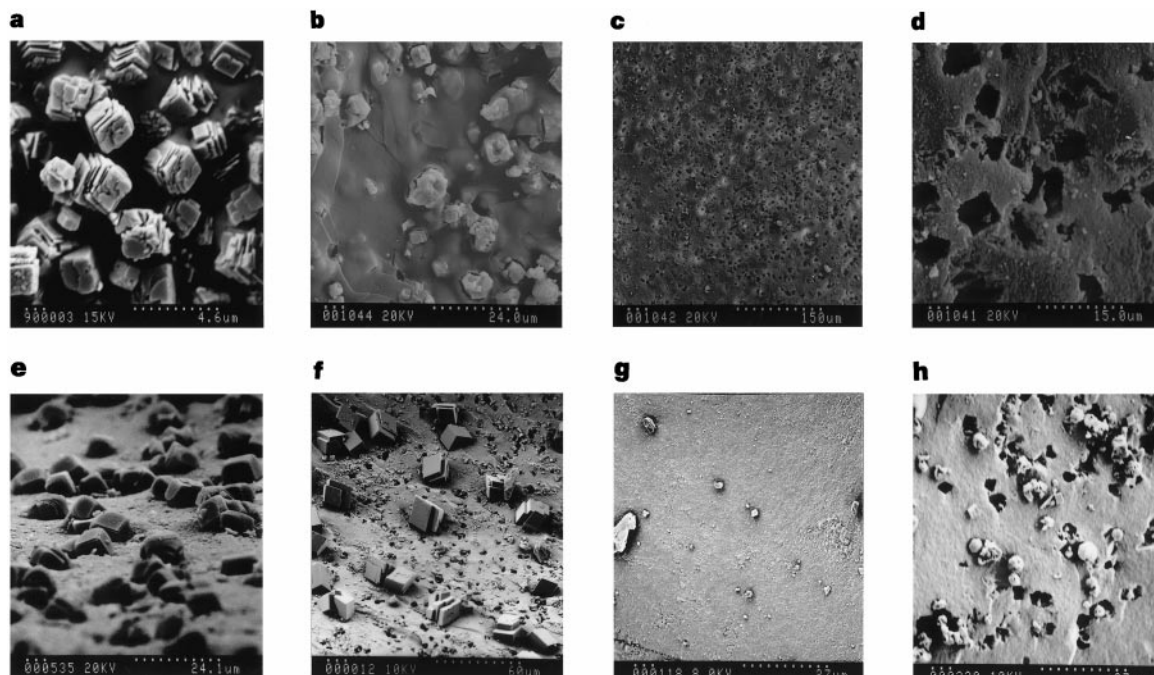


Figure 2 Nucleation of calcite at the polymer surface. Scanning electron micrographs (Hitachi S570) showing: **a**, template calcite crystals; **b**, Calcite-imprinted polymer (**PI-1**) as obtained; **c**, **d**, **PI-1** after HCl/CH₃OH wash; **e**, **f**, nucleation of calcite crystals on **PI-1**; **g**, **h**, control polymers **CP-1** and **CP-2** respectively, after immersion in supersaturated CaCO₃(aq.). Crystallization conditions: **a**, CaCl₂ (0.25 M), Na₂CO₃ (0.25 M); **e-h**, CaCl₂ (1.0 mM), Na₂CO₃ (0.8 mM). Crystallization studies were performed following the same general method, using triple-distilled water and the highest purity reagents available.

Crushed polymer (0.1 g) was placed in a bottle, and a stock solution of Na₂CO₃ (at 25 °C or 95 °C) was added, followed by an equal volume of CaCl₂ solution at the same temperature. The vessel was sealed (25 °C experiments) and the contents agitated for 24 h before filtration (0.2-μm membrane) and thorough washing of the polymer surface with water to remove weakly attached crystals. Polymer concentration was 10 mg ml⁻¹, except **f-h** (1.0 mg ml⁻¹). All glassware was washed with HCl (1 M, 3 × 50 ml), and water (3 × 50 ml) before use.

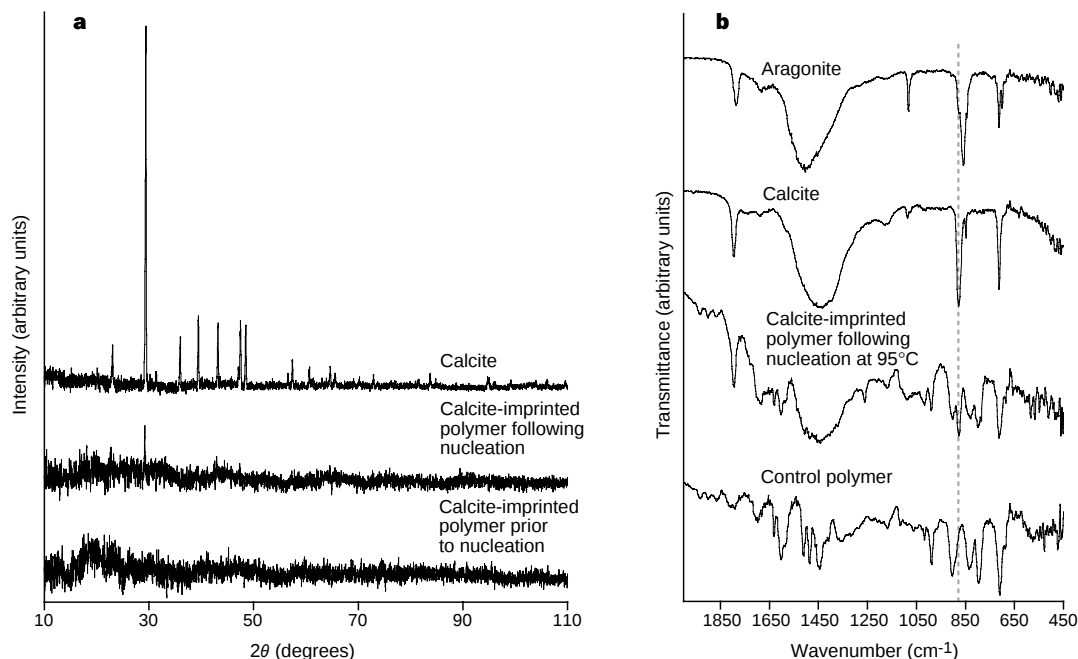


Figure 3 Characterization of CaCO₃ crystal polymorphs. **a**, Powder X-ray diffraction patterns of calcite crystals, imprinted polymer **PI-1** following incubation with supersaturated CaCO₃ solution, and the same polymer before nucleation but after removal of the template crystals. **b**, FT-IR spectra of aragonite and calcite crystals (showing characteristic bands at 858 cm⁻¹ and 877 cm⁻¹, respectively), imprinted polymer **PI-1** following incubation at 95 °C with super-

saturated CaCO₃ solution and washing (crystal-rich fraction, produced by grinding and sedimentation in chloroform), and similarly treated control polymer. X-ray diffraction patterns were recorded using Cu Kα radiation on a Inel Series 3000 CPS-120 X-ray diffractometer fitted with a Philips C-Tech tube. FT-IR spectra were obtained from KBr dispersions on a Perkin Elmer Series 1600 FT-IR spectrometer by the diffuse reflectance technique.

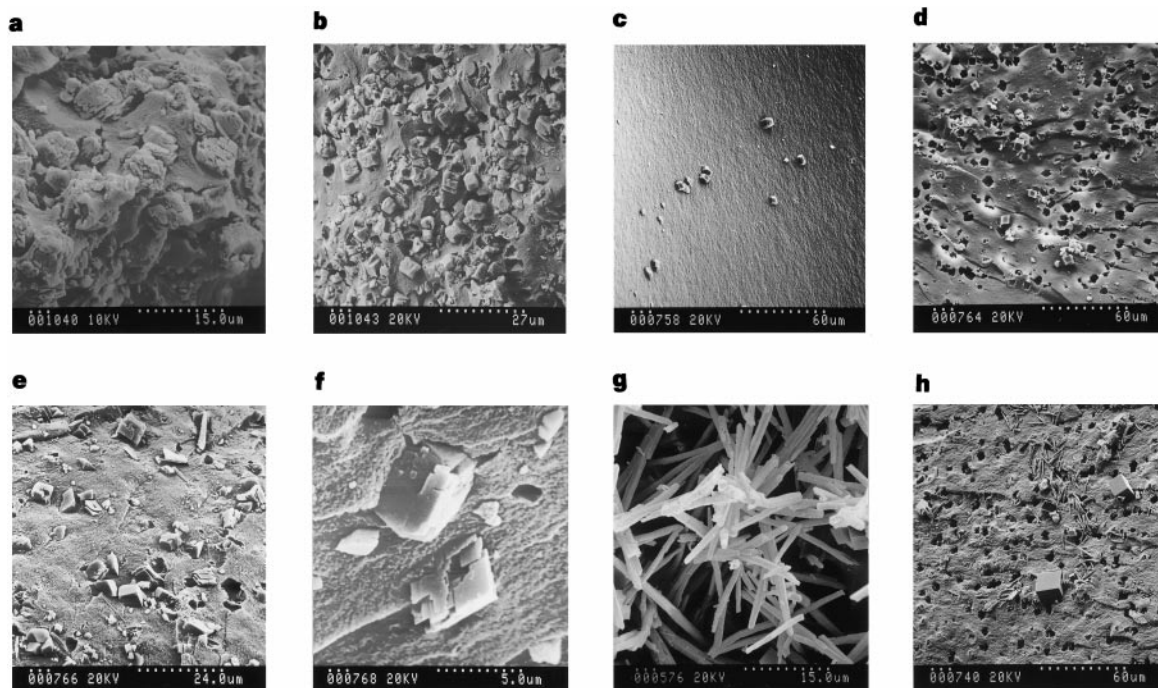


Figure 4 Calcite and aragonite nucleation in the presence of imprinted polymer. **a**, **b**, Nucleation of calcite on **PI-1** at 25 °C, **c**, **d**, Control polymers **CP-1** and **CP-2** respectively, after immersion in supersaturated CaCO_3 at 25 °C. **e**, **f**, Calcite on **PI-1** after crystallization at 95 °C. **g**, Aragonite crystals from solution at 95 °C (calcite content <4% as determined by powder X-ray diffraction). **h**, Control polymer **CP-2**

after crystallization at 95 °C. Initial concentrations of CaCl_2 and Na_2CO_3 were: **a**, 1.0 mM and 0.8 mM respectively; **b**–**h**, 1.35 mM and 1.35 mM except for **g** where the concentration of both substances was 10 mM. The crystallization mixtures **a**–**d** contained CH_3CN (10 vol.%).

The distribution of functional monomer within the polymer matrix was then assessed using a standard sodium hydroxide titration. We reasoned that if the incorporation of **1** in the imprinted sites occurred according to the mechanism outlined in Fig. 1, there should be a significant difference in the number of accessible reactive carboxyl groups between the polymer prepared in the presence (imprinted) and absence of template crystals. This was shown to be the case with $39 \pm 12 \mu\text{mol}$ and less than $10 \mu\text{mol}$ of carboxylic acid per gram of polymer found for the imprinted and control materials respectively out of a total of $79 \mu\text{mol}$, the remainder being inaccessible. The increased availability of carboxyl groups in the imprinted polymer should arise from their presence at the surface of the imprint sites. Indeed the difference is in moderately good agreement with the quantity of **1** found ($18 \mu\text{mol}$, by GC analysis) to bind to the template crystals in separate experiments performed in chloroform.

We also found that the distribution of calcite crystals grown on the surface of imprinted polymers varied markedly under our initial conditions, and many bare patches, similar to those of **CP-2** (Fig. 2h), were observed. This was attributed to the poor wettability of DVB polymers in water. In order to overcome the problem, the nucleation conditions were adjusted by the inclusion of 10 vol.% acetonitrile which was shown in separate experiments not to interfere with the crystallization of calcite. Under the modified protocol, the efficiency of nucleation was dramatically improved: up to 50% of sites were 'occupied' with crystals, as determined by visual inspection of random samples. In contrast, the number of crystals on the surface of control polymers, and imprinted polymers esterified with MeOH (using $p\text{-CH}_3\text{C}_6\text{H}_4\text{SO}_3\text{H}$ as catalyst) to block the carboxyl group before nucleation, remained negligible. As the modified protocol resulted in a smaller average crystal size (Fig. 4a), the concentration of ions was increased to 1.35 mM in later experiments to promote the growth of larger crystals (Fig. 4b).

The crystals grown under the more efficient nucleation conditions had a rather 'terraced' appearance (Fig. 4a, b) bearing greater

resemblance to the overall morphology of the template crystals (Fig. 2a) than observed in the earlier experiments. In the absence of an imprinted polymer matrix, the higher ion concentration resulted in the formation of regular, smooth-sided calcite rhombohedra. (The template crystals were formed at an even higher ion concentration of 0.25 M, which results in more pronounced overgrowth.) The terraced appearance of the polymer-grown crystals may have resulted from nucleation taking place concurrently at a number of points in the imprint site, followed by rapid growth in these areas until the individual crystals merge. Regular, sharp-edged rhombohedra, which presumably formed in solution and escaped the washing procedure, were occasionally seen on the surface of non-imprinted polymers (Fig. 4c, d). Their morphology is in contrast to that of the crystals formed in the imprint sites. We note that when the polymers were imprinted with better-quality rhombohedral calcite obtained at 5 mM, the crystals in the imprinted sites were found to be less 'kinked' in appearance, resembling the template material used.

Unambiguous evidence that the calcite polymorph was formed in the imprinted polymers is obtained from our FT-IR and powder XRD studies. Calcite has a strong reflection in the powder diffraction pattern at $2\theta = 29.36^\circ$, $d = 3.035 \text{ \AA}$ (Fig. 3a), and shows characteristic bands in the infrared spectrum at 1,795, 877 and 712 cm^{-1} (Fig. 3b). Imprinted polymer **PI-1** shows only amorphous scattering in XRD studies, whereas following nucleation, the characteristic calcite reflection can clearly be seen (Fig. 3a). Quantitative analysis confirmed that the calcite content of the crystals formed exceeded 96%, as determined by IR and XRD.

Having established that imprinted polymers were capable of nucleating calcite, we then investigated whether such polymers could be used to alter the course of crystallization. Experiments were repeated at 95 °C in pure water (acetonitrile could not be used as an aid to wetting at this temperature), conditions which strongly favour the formation of aragonite²¹. It is evident from Fig. 4e, f that even under these conditions characteristic calcite crystals still grew

in the imprinted sites, whereas predominantly aragonite was formed in solution, as expected. In some cases an occasional needle-like crystal of aragonite could be seen, which had probably been deposited onto the polymer particle during filtration and drying (compare Fig. 4e and g). More significantly, examination of the surface of control polymer CP-2 (Fig. 4h) revealed the presence of deposited aragonite needles as well as a few sharp-edged 'solution-type' calcite crystals. A crystal-enriched fraction from PI-1, prepared by grinding and sedimentation in chloroform, was shown conclusively to contain calcite, both by FT-IR (Fig. 3b) and XRD (not shown).

We attempted to produce aragonite-specific polymers using the same monomer **1** and the needles shown in Fig. 4g as template. But the results of this experiment were inconclusive, as the polymers did not nucleate as well as those imprinted with calcite and neither FT-IR nor powder XRD were sensitive enough to quantify the ratio of aragonite to calcite on the polymer surface. It is not clear why the imprinting of calcite should result in much more efficient nucleation than the corresponding aragonite imprints. However, as the arrangement of ions on the crystal planes of each polymorph are different, it is reasonable to expect that a functionality which is suitable for creating complementary surfaces in one case may not necessarily be useful for another type of crystal. We note that the use of acrylic and methacrylic acids in place of the monomer **1** resulted in imprinted polymers with inferior nucleating abilities. The reason for this is not clear, but may be due to the presence of the flexible spacer which separates the acid head group from the polymerizable double bond of **1**, allowing this monomer to match more easily the spacing of ions on the crystal surface and for this ordering to be preserved throughout the entire imprinting process. Better understanding of these processes may enable the design of devices for generating crystals of the desired type, size or shape for applications in materials science. In addition, the ability to generate surfaces specific to crystals of an enantiomerically pure organic compound—to promote its crystallization from a racemic mixture—may lead to the introduction of new separation methods in the pharmaceutical and fine-chemical industries. □

Received 10 August 1998; accepted 13 January 1999.

- Mann, S. Molecular recognition in biomineralization. *Nature* **332**, 119–124 (1988).
- Berman, A. *et al.* Intercalation of sea-urchin proteins in calcite—study of a crystalline composite material. *Science* **250**, 664–667 (1990).
- Mann, S. Molecular tectonics in biomineralization and biomimetic materials chemistry. *Nature* **365**, 499–505 (1993).
- Falini, G., Albeck, S., Weiner, S. & Addadi, L. Control of aragonite or calcite polymorphism by mollusk shell macromolecules. *Science* **271**, 67–69 (1996).
- Mann, S., Heywood, B. R., Rajam, S. & Birchall, J. D. Controlled crystallization of CaCO₃ under stearic-acid monolayers. *Nature* **334**, 692–695 (1988).
- Bianconi, P. A., Lin, J. & Strzelecki, A. R. Crystallization of an inorganic phase controlled by a polymer matrix. *Nature* **349**, 315–317 (1991).
- Bunker, B. C. *et al.* Ceramic thin-film formation on functionalized interfaces through biomimetic processing. *Science* **264**, 48–55 (1994).
- Walsh, D. & Mann, S. Fabrication of hollow porous shells of calcium carbonate from self-organizing media. *Nature* **377**, 320–323 (1995).
- Oliver, S., Kuperman, A., Coombs, N., Lough, A. & Ozin, G. A. Lamellar aluminophosphates with surface patterns that mimic diatom and radiolarian microstructures. *Nature* **378**, 47–50 (1995).
- Berman, A. *et al.* Total alignment of calcite at acidic polydiacetylene films—cooperativity at the organic-inorganic interface. *Science* **269**, 515–518 (1995).
- Mann, S. & Ozin, G. A. Synthesis of inorganic materials with complex form. *Nature* **382**, 313–318 (1996).
- Ozin, G. A. Morphogenesis of biomineral and morphosynthesis of biomimetic forms. *Acc. Chem. Res.* **30**, 17–27 (1997).
- Yang, H., Coombs, N. & Ozin, G. A. Morphogenesis of shapes and surface patterns in mesoporous silica. *Nature* **386**, 692–695 (1997).
- Küther, J., Seshadri, R., Knoll, W. & Tremel, W. Templated growth of calcite, vaterite and aragonite crystals on self-assembled monolayers of substituted alkythiols on gold. *J. Mater. Chem.* **8**, 641–650 (1998).
- Ozin, G. A. *et al.* Bone mimetics: A composite of hydroxyapatite and calcium dodecylphosphate lamellar phase. *J. Mater. Chem.* **7**, 1601–1607 (1997).
- Heywood, B. R. & Mann, S. Template-directed nucleation and growth of inorganic materials. *Adv. Mater.* **6**, 9–20 (1994).
- Wulff, G. Molecular imprinting in cross-linked materials with the aid of molecular templates—a way towards artificial antibodies. *Angew. Chem. Int. Edn Engl.* **34**, 1812–1832 (1995).
- Mosbach, K. & Ramström, O. The emerging technique of molecular imprinting and its future impact on biotechnology. *Bio/Technology* **14**, 163–170 (1996).
- Shea, K. J. Molecular imprinting of synthetic network polymers: The *de novo* synthesis of macro-molecular binding and catalytic sites. *Trends Polym. Sci.* **2**, 166–173 (1994).
- Vulfson, E. N., Alexander, C. & Whitcombe, M. J. Assembling the molecular cast. *Chem. Brit.* **33**, 23–26 (1997).

- De Keyser, W. L. & Degeldre, L. Contribution à l'étude de la formation de la calcite, aragonite et vaterite. *Bull. Soc. Chim. Belg.* **59**, 40–71 (1950).
- Reddy, M. M. & Nancollas, G. H. The crystallization of calcium carbonate IV. The effect of magnesium, strontium and sulfate ions. *J. Cryst. Growth* **35**, 33–38 (1976).
- Batz, H.-G. & Koldehoff, J. Monomere und polymere succinimidoester von ω -methacryloylamino-säuren, ihre darstellung und ihre reaktion mit aminen. *Makromol. Chem.* **177**, 683–686 (1976).

Acknowledgements. We thank A. Jahans for the X-ray crystallographic studies. This work was supported by BBSRC and Unilever plc.

Correspondence and requests for materials should be addressed to M.J.W. (e-mail: michael.whitcombe@bbsrc.ac.uk).

Ubiquity of quasi-horizontal layers in the troposphere

Reginald E. Newell*, Valerie Thouret†, John Y. N. Cho*, Patrick Stoller‡, Alain Marenco† & Herman G. Smit§

* Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Laboratoire d'Aérodynamique (CNRS-URA OMP 0354), F-31400 Toulouse, France

§ Research Center Jülich, Institute for Chemistry of the Polluted Atmosphere (ICG-2), PO Box 1913, 52425 Jülich, Germany

Fine laminar structures in the atmosphere have been described previously^{1–9}, but their characterization has been limited. The modern global coverage of aircraft flights offers an opportunity to provide such a characterization, and examine the ubiquity of such structures, in space and time. Research aircraft measuring vertical profiles of atmospheric chemical constituents frequently discern quasi-horizontal atmospheric layers with mean thicknesses of the order of 1 km and mean altitudes between 5 and 7 km (refs 10–12). These layers can be characterized and categorized by various combinations of ozone, water vapour, carbon monoxide and methane deviations from background profiles. Five commercial aircraft have been recently equipped to measure water vapour and ozone concentrations, and automatically collect vertical profile information on landing and take-off (refs 13–15). Here we synthesize measurements from both research and commercial flights and demonstrate the ubiquity in space and time of four layer types (as categorized by their chemical signatures). Up to one-fifth of the lowest 12 km of the atmosphere is occupied by such layers. We suggest that this universality reflects basic characteristics of the atmosphere hitherto unexplored, with potential implications for present understanding of a wide variety of dynamic and chemical atmospheric processes.

Atmospheric vertical structure has traditionally been measured with temperature and water-vapour sensors on balloon-borne radiosondes rising at about 5 m s⁻¹. The balloon is tracked to provide wind data. Although data can be collected at 6-s intervals, they are generally only reported at height intervals of 1.5–2.5 km, so a climatology of smaller-scale atmospheric structure is not available. Research aircraft measure many different atmospheric trace constituents while carrying out studies of ozone and other trace constituent budgets, providing a more precise identification of atmospheric layers. In addition, five Airbus A-340 aircraft in regular commercial service now measure ozone^{13,14} and water vapour^{14,15} continuously in flight, covering a substantial geographical area (the "MOZAIC" programme). Each time these aeroplanes take off and land an atmospheric vertical profile is traced out. We have used aircraft data from MOZAIC and three NASA missions: Pacific Exploratory Mission-West A (PEMA), in the western Pacific, September–October 1991; PEM-West B (PEMB), February–March 1994; and PEM-Tropics (PEMT), in the central and South

* Present address: Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA.

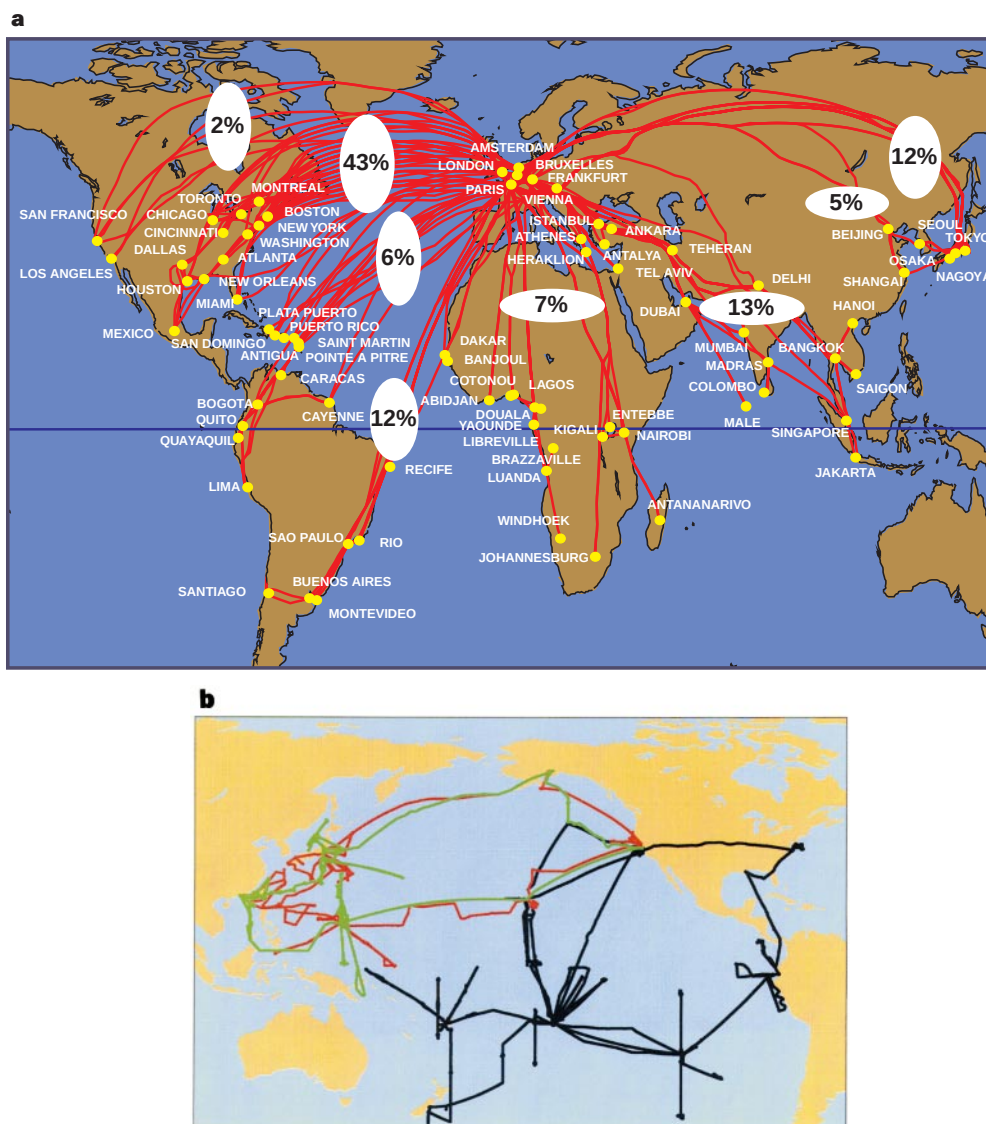


Figure 1 Maps of regions sampled. **a**, MOZAIC flight track and percentage of total flights (7,500) in indicated regions, September 1994–December 1997. Vertical profiles were only collected on take-off and landing. **b**, PEM flight tracks for 64

flights. Red, PEMA; green, PEMB; black, PEM tropics. Vertical profiles were collected throughout each flight.

Pacific, September–October 1996. Descriptions of PEMA, PEMB and PEMT with flight plans, instrument information and scientific findings, including layer examples and statistical summaries, are available^{10–12,16–18}; MOZAIC and PEM coverages are shown in Fig. 1. MOZAIC is clearly more extensive; three full years of data are used here and the programme continues, while PEM was limited to the six months listed above. Comparison with the MOZAIC data thus provides a broader perspective on small-scale structure.

For each constituent profile, we estimated a background using a mode-based algorithm¹². We characterized layers by calculating deviations from this background profile. The points with deviations of ± 10 parts per billion by volume (p.p.b.v.) for O_3 and $\pm 5\%$ for water vapour as relative humidity were defined as layers. Ozone was examined first, then water vapour; if both exceeded these limits, we defined a layer. Layer thickness was determined by finding the nearest points where the ozone deviation was below 5 p.p.b.v. in magnitude. For the research aircraft data, CO and CH_4 were also included, with deviations of ± 3 p.p.b.v., to define a second set of layers for which all four criteria were exceeded¹². The much larger MOZAIC data sample permitted exploring variations in the magnitudes of the criteria for a layer from 5 p.p.b.v. to 20 p.p.b.v. for ozone and 5% to 20% for water vapour. To identify the layers,

deviations from background are henceforth indicated by plus and minus signs following the chemical symbol: O_3+/H_2O- labels a layer characterized by a positive deviation from the background ozone profile coupled with a negative deviation from the background water-vapour profile. Results (Table 1) show increases in the number of layers for the smaller criteria. Average thicknesses of the layers are between 0.5 and 1.3 km and average altitudes are between 5.5 and 6.6 km.

A preliminary interpretation of the four combinations, based on the research aircraft data, is: O_3+/H_2O+ , continental pollution; O_3+/H_2O- , stratospheric air or pollution; O_3-/H_2O+ , convection from the boundary layer; and O_3-/H_2O- , subsiding air originally raised in deep convection over oceans. The properties of the different types vary with the anomaly criteria selected. The average altitude decreases as the criterion for water-vapour anomaly increases, while larger ozone anomalies yield higher altitudes for O_3+ . Table 2a shows the universality in percentage distribution of the four layer types between the research aircraft and MOZAIC samples for $O_3 \pm 10$ p.p.b.v. and relative humidity $\pm 5\%$, despite geographical coverage and seasonal differences. The O_3+/H_2O- layer type is always the most abundant. For the same criteria, Table 2b shows that the mean heights of the different layer types

vary over a small range, while the thicknesses in PEMT are the smallest overall, probably a result of the higher data resolution during PEMT¹². From the total layer number for each sample set in Table 2b, their average thickness, and the vertical span sampled (Table 2c), we calculate that about 15% of the atmosphere is occupied by layers. For MOZAIC, this value diminishes to 7% if the large O₃ anomaly criterion (20 p.p.b.v.) is used, or to 8% if the large H₂O anomaly criterion (20%) is used. Use of four layer criteria enables the O₃+H₂O- layers to be subdivided into those with stratospheric origins and those from pollution, as the former always have low concentrations of CO and usually low CH₄. However, the subdivision is complicated by the fact that high stability often accompanies stratospheric air in the troposphere and tends to trap pollutants. The conclusion that this type of layer is always the most abundant could only have been made with the extensive time and space coverage of MOZAIC; it reflects the fact that the transfer of stratospheric air into the troposphere is important in the atmospheric general circulation.

Danielsen long ago stressed the laminar structure of the atmosphere¹; he was concerned with how radioactive debris injected into the stratosphere moved back into the troposphere². He pointed out that atmospheric fine structure was omitted in normal reporting of balloon-borne radiosonde data and gave examples of how the structure could be recovered by reprocessing. Corby³ and others⁴⁻⁶ also used radiosonde data to study tropospheric fine structure, while Sawyer⁷ extended the work to the lower stratosphere using a radar-sonde theodolite. Fine structure of the 17–70 km region was also examined by tracking rocket-launched falling spheres by radar⁸. More recently, powerful clear-air radars have enabled wind and turbulence layers to be measured throughout the atmosphere⁹. Except for Danielsen the general interpretation of this work was that the layers were evidence of gravity waves. These waves are associated with wind velocity changes and some are thought to be orographically forced. But in the layers revealed by trace constituent measurements, there are often no wind velocity signatures evident¹¹. Furthermore, the extensive MOZAIC observations show such structure over both land and sea, and frequently in regions where orographic forcing is absent.

Layers identified as continental pollution and marine boundary layer convection move upwards near their source regions, whereas stratospheric layers and clean marine air generally subside. Subsidence is particularly effective for layers drier than the background environ-

Table 1 MOZAIC layer data

	Minimum H ₂ O deviation, 5%			
	Min. O ₃ dev.			
	5 p.p.b.v.	10 p.p.b.v.	15 p.p.b.v.	20 p.p.b.v.
Number				
O ₃ +H ₂ O+	8,658	3,341	1,676	977
O ₃ +H ₂ O-	20,771	11,288	6,703	4,327
O ₃ -H ₂ O+	10,634	3,751	1,411	672
O ₃ -H ₂ O-	11,284	4,000	1,457	657
Thickness (m)				
O ₃ +H ₂ O+	475	679	927	1,083
O ₃ +H ₂ O-	689	867	1,135	1,358
O ₃ -H ₂ O+	515	779	1,113	1,313
O ₃ -H ₂ O-	494	736	1,072	1,353
Height (m)				
O ₃ +H ₂ O+	5,952	6,321	6,472	6,653
O ₃ +H ₂ O-	5,543	5,735	5,888	6,042
O ₃ -H ₂ O+	5,753	5,872	5,830	5,657
O ₃ -H ₂ O-	5,855	5,994	6,100	6,093
Minimum O ₃ deviation, 10 p.p.b.v.				
	Min. H ₂ O dev.			
	5%	10%	15%	20%
Number				
O ₃ +H ₂ O+	3,341	2,323	1,605	1,075
O ₃ +H ₂ O-	11,288	8,846	6,656	5,266
O ₃ -H ₂ O+	3,751	2,776	2,009	1,408
O ₃ -H ₂ O-	4,000	2,706	1,829	1,343
Thickness (m)				
O ₃ +H ₂ O+	679	710	739	719
O ₃ +H ₂ O-	867	907	940	969
O ₃ -H ₂ O+	779	795	804	820
O ₃ -H ₂ O-	736	743	755	751
Height (m)				
O ₃ +H ₂ O+	6,321	6,142	5,965	5,805
O ₃ +H ₂ O-	5,735	5,637	5,571	5,478
O ₃ -H ₂ O+	5,872	5,730	5,587	5,512
O ₃ -H ₂ O-	5,994	5,900	5,837	5,780

These data are for August 1994–December 1997.

ment, as radiative effects give relative cooling near the layer base which tends to move the base downwards independently of any large-scale subsidence¹². Radiative heating rate calculations for an O₃-H₂O+ layer in the upper troposphere, using a standard radiative transfer program¹⁹, give heating near the base and cooling near the top (Fig. 2) which yields a tendency towards destabilization. Such layers would be likely regions for the occurrence of clear air turbulence.

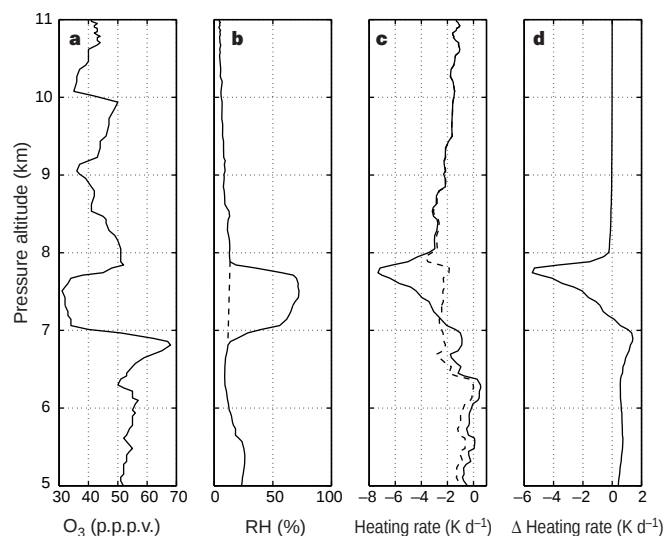


Figure 2 Radiative heating rate profiles for an O₃-H₂O+ layer versus a simulated background, both without clouds. The data were obtained during a MOZAIC flight landing in Osaka on 12 November 1996, 00:25 UT. **a**, Ozone profile. **b**, Measured (solid line) and interpolated background (dashed) relative humidity (RH) profiles.

The latter simulates a layerless situation. **c**, Calculated heating rate profiles for the measured RH (solid) and the interpolated layerless situation (dashed). **d**, Heating rate for the actual data minus the heating rate for the simulated background.

Table 2 Occurrence and characteristics of atmospheric layers

(a) Percentages of layer types for MOZAIC* and PEM

	O ₃ + /H ₂ O+	O ₃ + /H ₂ O-	O ₃ - /H ₂ O+	O ₃ - /H ₂ O-
PEMA	14	54	15	18
PEMB	11	54	17	18
PEMT	12	53	19	17
MOZAIC	15	50	17	18

(b) Number, thickness and height for layer types

	O ₃ + /H ₂ O+			O ₃ + /H ₂ O-			O ₃ - /H ₂ O+			O ₃ - /H ₂ O-		
	No.	Th.†	H.†	No.	Th.†	H.†	No.	Th.†	H.†	No.	Th.†	H.†
PEMA	11	0.68	5.7	43	0.76	5.2	12	0.87	5.1	14	0.86	6.5
PEMB	8	0.59	6.5	38	0.99	6.0	12	0.64	6.0	13	0.49	6.4
PEMT	27	0.45	6.0	120	0.71	5.2	43	0.47	5.4	38	0.45	6.1
MOZAIC	3,341	0.67	6.3	11,288	0.86	5.7	3,751	0.78	5.8	4,000	0.74	5.9

(c) Percentage of atmosphere occupied by layers

	Profiles‡ (km)	Avg. thickness (km)	Total no. of layers	% occupied
PEMA	439	0.78	80	14
PEMB	388	0.79	71	14
PEMT	655	0.59	228	20
MOZAIC	105,498	0.79	22,380	17

* From Table 1, upper section, O₃ ± 10 p.p.b.v.

† Thickness (Th.) and height (H.) in km.

‡ Total vertical distance sampled.

The relative invariance of percentages in Table 2a may reflect the overall mass balance that must occur. While the layers measured are mostly natural, some of the layers induced by biomass burning are from fires set by humans. Thus some anthropogenic effect on the buoyancy over the continents exists; this will be reflected elsewhere in more subsidence.

Atmospheric layers apparently play a significant role in the large-scale general circulation, in addition to the stratosphere–troposphere exchange noted above. Much upward motion in the atmosphere occurs via large-scale convergence in the lower atmosphere, followed by deep convection. This often culminates in a layer formed as the convected air spreads laterally in the upper troposphere. Upward motion also occurs in cyclones, and is often itself a quasi-horizontal layer-type process. Clouds are manifestations of these upward motions. The sinking motion which must accompany these processes is not always so evident. Dynamical and chemical numerical models used for weather forecasting, chemical-species time evolution and studies of climatic fluctuations should use a vertical resolution that is sufficient to define layers. Such resolution would be about 0.2 km or better, according to our observations—much higher than present-day models. The desired resolution for chemical models, in which highly nonlinear relationships are involved in ozone production²⁰, may be higher²¹. To accompany the enhanced resolution, the models will need to contain a new or modified formalism to include laminae.

When layers are included in atmospheric diagnostic and modelling studies, there are consequences for the mass, momentum and energy budgets. For example, layers of stratospheric air can be followed in the troposphere and their contribution to mass and trace-constituent exchange between the regions estimated. Another example is that the radiative cooling from layers rich in water vapour in the upper troposphere is substantially different from that of an upper troposphere with uniformly mixed water vapour (Fig. 2). The vertical gradients of cooling tend to produce instability which will influence the vertical energy and mass transport. A practical consequence is that clear air turbulence (CAT) could be triggered in such regions. Studies could be made of the occurrence of CAT near water vapour layers at high altitude such as those observed by MOZAIC; if there turned out to be an association, the layers could eventually be observed regularly with modified satellite limb scanning instruments. These have been used to delineate the upper-tropospheric water-vapour field and give results consistent

with the large-scale atmospheric divergence field²². To detect the smaller-scale circulation features, the modifications would need to improve the vertical and horizontal resolution. □

Received 21 September 1998; accepted 7 January 1999.

- Danielsen, E. F. The laminar structure of the atmosphere and its relation to the concept of a tropopause. *Arch. Meteorol. Geophys. Bioklimatol.* **11**, 293–332 (1959).
- Danielsen, E. F. Stratospheric–tropospheric exchange based on ozone and potential vorticity. *J. Atmos. Sci.* **25**, 502–518 (1968).
- Corby, G. A. A preliminary study of atmospheric waves using radiosonde data. *Q. J. R. Meteorol. Soc.* **83**, 49–60 (1957).
- Reid, S. J. An observational study of lee waves using radiosonde data. *Tellus* **24**, 593–596 (1982).
- Shutts, G. J., Kitchen, M. & Hoare, P. H. A large amplitude gravity wave in the lower stratosphere detected by radiosonde. *Q. J. R. Meteorol. Soc.* **114**, 579–594 (1988).
- Bethan, S., Vaughan, G. & Reid, S. J. A comparison of ozone and thermal tropopause heights and the impact of tropopause definition on quantifying the ozone content of the troposphere. *Q. J. R. Meteorol. Soc.* **122**, 929–944 (1996).
- Sawyer, J. S. Quasi-periodic wind variations with height in the lower stratosphere. *Q. J. R. Meteorol. Soc.* **87**, 24–33 (1961).
- Newell, R. E., Mahoney, J. R. & Lenhard, R. W. Jr A pilot study of small-scale wind variations in the stratosphere and mesosphere. *Q. J. R. Meteorol. Soc.* **92**, 41–54 (1966).
- Gage, K. S. & Balsley, B. B. MST radar studies of wind and turbulence in the middle atmosphere. *J. Atmos. Terr. Phys.* **46**, 739–753 (1984).
- Newell, R. E. *et al.* Vertical fine-scale atmospheric structure measured from NASA DC-8 during PEM-West A. *J. Geophys. Res.* **101**, 1943–1960 (1996).
- Wu, Z.-X. *et al.* Atmospheric layers measured from the NASA DC-8 during PEM-West B and comparison with PEM-West A. *J. Geophys. Res.* **102**, 28353–28365 (1997).
- Stoller, P. *et al.* Measurements of atmospheric layers from the NASA DC-8 and P-3B aircraft during PEM-Tropics A. *J. Geophys. Res.* (in the press).
- Thouret, V., Marenco, A., Logan, J. A., Nédélec, P. & Grouhel, C. Comparisons of ozone measurements from the MOZAIC airborne program and the ozone sounding network at eight locations. *J. Geophys. Res.* **103**, 25695–25720 (1998).
- Marenco, A. *et al.* Measurement of ozone and water vapour by Airbus in-service aircraft: The MOZAIC airborne program, an overview. *J. Geophys. Res.* **103**, 25631–25642 (1998).
- Helten, M. *et al.* Calibration and performance of automatic compact instrumentation for the measurement of relative humidity from passenger aircraft. *J. Geophys. Res.* **103**, 25643–25652 (1998).
- Hoell, J. M. *et al.* Pacific exploratory mission-west A (PEM-West A): September–October 1991. *J. Geophys. Res.* **101**, 1641–1653 (1996).
- Hoell, J. M. *et al.* The Pacific exploratory mission-west phase B: February–March, 1994. *J. Geophys. Res.* **102**, 28223–28239 (1997).
- Hoell, J. M. *et al.* The Pacific exploratory mission in the tropical Pacific: PEM-Tropics A, August–September, 1996. *J. Geophys. Res.* (in the press).
- Hoffman, R. N. A Computer Program which Calculates Radiative Fluxes and Heating Rates in Model Atmospheres (Sci. Rep. No. 4, Dept Meteorol. Phys. Oceanogr., MIT, Cambridge, MA, 1981).
- Brasseur, G. P., Müller, J. F. & Granier, C. Atmospheric impact of NO_x emissions by subsonic aircraft: A three dimensional model study. *J. Geophys. Res.* **101**, 1423–1428 (1996).
- Edouard, S., Legras, B., Lefèvre, F. & Eymard, R. The effect of small-scale inhomogeneities on ozone depletion in the Arctic. *Nature* **384**, 444–447 (1996).
- Newell, R. E., Zhu, Y., Browell, E. V., Read, W. G. & Waters, J. W. Walker circulation and tropical upper tropospheric water vapor. *J. Geophys. Res.* **101**, 1961–1974 (1996).

Acknowledgements. We thank our PEM and MOZAIC colleagues for access to their data, Air France, Lufthansa, Austrian Airlines and Sabena for carrying the MOZAIC equipment, and the NASA GTE, ACMAF, and GSFC Research Assistantship programs for their support.

Correspondence and requests for materials should be addressed to R.E.N. (e-mail: newell@newell1.mit.edu).

Oceanic forcing of the wintertime North Atlantic Oscillation and European climate

M. J. Rodwell*, D. P. Rowell* & C. K. Folland*

* Hadley Centre for Climate Prediction and Research, UK Meteorological Office, London Road, Bracknell RG12 2SY, UK

The weather over the North Atlantic Ocean, particularly in winter, is often characterized by strong eastward air-flow between the 'Icelandic low' and the 'Azores high', and by a 'stormtrack' of weather systems which move towards western Europe. The North Atlantic Oscillation—an index of which can be defined as the difference in atmospheric pressure at sea level between the Azores and Iceland—is an important mode of variability in the global atmosphere^{1,2} and is intimately related to the position and strength of the North Atlantic stormtrack owing to dynamic processes internal to the atmosphere^{3,4}. Here we use a general circulation model of the atmosphere to investigate the ocean's role in forcing North Atlantic and European climate. Our simulations indicate that much of the multiannual to multidecadal variability of the winter North Atlantic Oscillation over the past half century may be reconstructed from a knowledge of North Atlantic sea surface temperature. We argue that sea surface temperature characteristics are 'communicated' to the atmosphere through evaporation, precipitation and atmospheric-heating processes, leading to changes in temperature, precipitation and storminess over Europe. As it has recently been proposed that there may be significant multiannual predictability of North Atlantic sea surface temperature patterns⁵, our results are encouraging for the prediction of European winter climate up to several years in advance.

It has been long recognized that fluctuations in sea surface temperature (SST) and the strength of the North Atlantic Oscillation (NAO) in the North Atlantic are related⁶. For example, stronger eastward flow generally increases evaporation, thereby cooling SST^{6–9}. On the other hand, SST anomalies, including those in middle latitudes, have been thought to be partly responsible for changes in atmospheric circulation over the North Atlantic and Europe in winter or spring^{10–13}. A mechanism has been proposed^{14,15} that involves a positive feedback between the ocean and atmosphere. Other authors, however, find little evidence of positive feedbacks^{16,17}.

The global atmospheric model used in this study is the HadAM2b version of the UK Meteorological Office's Unified Model¹⁸. It has a horizontal grid resolution of $2.5^\circ \times 3.75^\circ$ in latitude and longitude with 19 hybrid σ , p levels in the vertical and a timestep of 30 min. The model incorporates prognostic cloud water and ice, has a mass-flux convection scheme with stability closure and uses mean orography. It was forced with observed SSTs and sea-ice extents taken from the GISST3.0 data set (N. Rayner, personal communication). An ensemble of six 128-year simulations were started from 1870, differing only in their initial atmospheric conditions. By taking averages of this ensemble, atmospheric 'noise' can be substantially reduced. Here, we discuss the period 1947–97 when oceanic data are thought to be most reliable. We focus on the December to February (DJF) season when, through strong heat fluxes at the surface and a deeper mixed layer in the ocean⁹, the atmosphere is likely to be most strongly linked to deeper ocean temperatures, which vary on the longer timescales of interest here.

Figure 1 shows that the model simulates well the observed decline

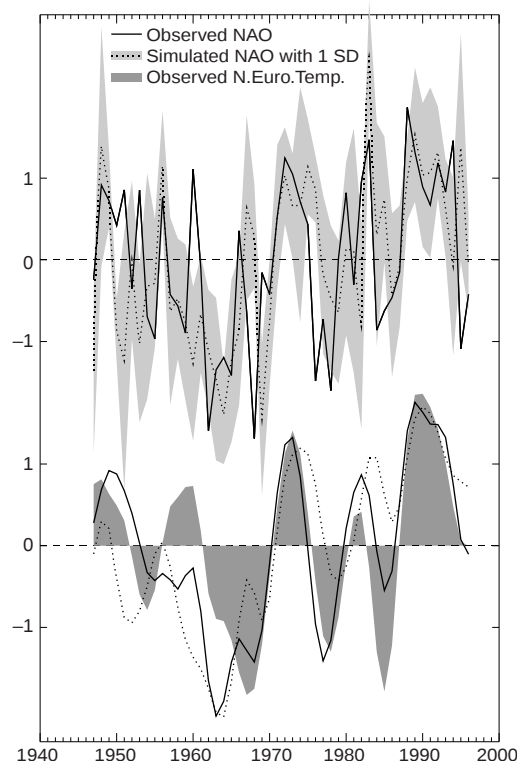


Figure 1 Time series of the North Atlantic Oscillation index. Observed (solid line) and modelled ensemble average (dotted line) winter NAO index, December to February 1947–97. The NAO index is calculated as the normalized difference in December to February mean sea-level pressure between the Azores ($26^\circ\text{W } 38^\circ\text{N}$) and Iceland ($23^\circ\text{W } 65^\circ\text{N}$). Observed data is taken from Ponta Delgada (Azores) and Stykkisholmur (Iceland). The shading in the upper graph shows ± 1 standard deviation about the ensemble mean, calculated from the normalized six model simulations for each individual year. The lower graph shows the NAO index time series after they have been filtered to pass variations with periods greater than 6.5 years (using a 1-2-1 binomial filter three times). Shading in the lower graph is the normalized filtered time series of observed North European surface temperatures (averaged over the box $5-50^\circ\text{E}$, $50-70^\circ\text{N}$). The year corresponds to December for each DJF season.

of the NAO to the 1960s, its subsequent rise to the early 1990s, and some of the recent smaller fall. Moreover, much of the multiannual variability and some interannual changes in the NAO have been captured. The correlation between the simulated and observed unfiltered curves is quite high at 0.41 (upper set of curves in Fig. 1) and significant at the 98% confidence level allowing for persistence. These curves show that, assuming SSTs are known, it may be possible to predict the correct sign of the winter NAO index in 2 out of 3 years. The sign of strong events (larger than one standard deviation) is correctly predicted 3 times out of 4. Each individual model simulation correlates more poorly with the observed NAO index than does the ensemble average, emphasizing the importance of making more than one model simulation. Internal variability in the real atmosphere still limits the accuracy of a forecast, although the observed NAO index does generally remain within ± 1 standard deviation of the ensemble mean (indicated by the shading in the upper graph in Fig. 1). When internal variability is reduced by filtering the data to include only timescales longer than 6.5 years (lower curves in Fig. 1), the correlation rises to 0.74 which is significant at the 99.9% level, again allowing for persistence. Thus multiannual to multidecadal variations in the NAO are considerably better simulated than interannual timescales. This study uses an atmospheric model,

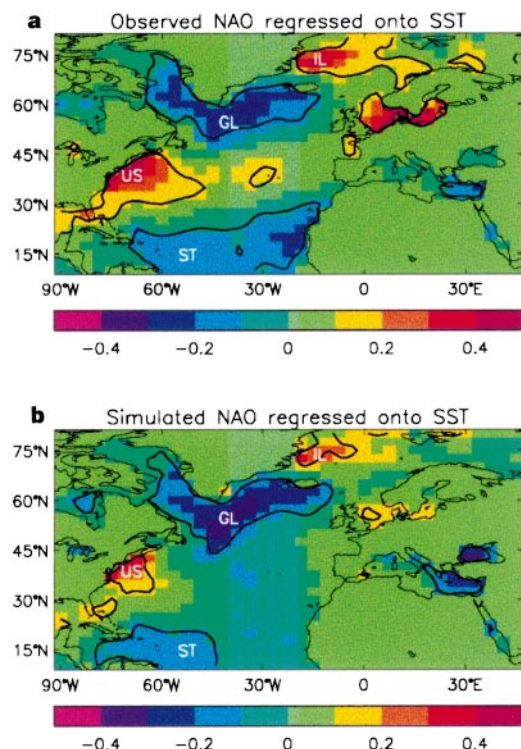


Figure 2 NAO/SST regressions. **a**, Regression of the 50-year time series of observed unfiltered NAO index, as in Fig. 1, onto observed North Atlantic SST. Units are in K and correspond to an NAO index anomaly of 1 standard deviation. **b**, As **a** but for a regression of the unfiltered model ensemble averaged NAO index onto North Atlantic SST. The areas within the black contours exceed the 95% confidence level of non-zero correlation using a 2-tailed *t*-test. Regions 'ST', 'US', 'GL' and 'IL' are explained in the text.

without feedback onto the ocean, and hence the skill of our NAO simulations must result from the prescribed SSTs and sea ice. We note that any direct effect of changes in greenhouse gases or stratospheric ozone on the NAO is not included in this simulation, although indirect effects through the ocean surface would be included.

The NAO has a substantial effect on temperature and rainfall in Europe and the eastern United States^{20–22}. The shading in the lower graph of Fig. 1 shows the filtered time series of observed North European surface temperatures. Their correlation with the observed NAO index is very high at 0.82 (the unfiltered correlation is 0.72), but the model does somewhat less well at capturing this link. The difference between the two 7-year winter periods December–February 1962–69 (low NAO index) and DJF 1987–94 (high NAO index) shows precipitation²³ increased by up to 30% in Britain and much of Scandinavia and decreased by 50% in southern Spain, Italy and Greece. Hence there are benefits to be gained, for instance for agriculture and the utilities²⁴, if skilful long-range forecasts of the NAO can be made.

Figure 2a shows the regression of the observed unfiltered NAO index onto SSTs for much of the North Atlantic region. The regression pattern, with negative centres in the subtropics (ST) and south of Greenland (GL) and positive centres off the east coast of the United States (US) and north of Iceland (IL), agrees well with previous results^{5,6,13,15,25}. Outside the region shown, correlations are not field-significant (that is, the small percentage of points with appreciable correlations could have occurred by chance), with no locally statistically significant correlations in the tropical and northern Pacific Ocean or the South Atlantic Ocean. Neither is there any evidence of locally statistically significant correlation when the NAO index is regressed onto winter mean sea-level pressure (MSLP)

outside the North Atlantic region. These results suggest that El Niño and any pan-Atlantic SST mode²⁶ have little direct relationship to the NAO in this season³ (unless in relatively rare circumstances). When the modelled ensemble mean NAO index is regressed onto the same SST data, a similar pattern emerges in the North Atlantic, Fig. 2b, suggesting that ocean to atmosphere forcing is important at this timescale and that the model correctly simulates the main mechanisms linking SSTs to the NAO, such as surface fluxes, atmospheric convection and any teleconnections within the North Atlantic region. Differences between Fig. 2a and b in the North Sea and Baltic and off the coast of northwest Africa may suggest that atmosphere to ocean forcing is dominant in these regions.

To confirm that North Atlantic SSTs do force the NAO, we use a North Atlantic SST pattern similar to Fig. 2a and b that is obtained from a canonical correlation analysis by Grötzner *et al.*¹⁵ (Fig. 3a). It is appropriately scaled to force 20-year idealized simulations with the model. Maximum values of the SST forcing are about $\pm 1^\circ\text{C}$. Figure 3b shows the difference in DJF MSLP between model simulations with positive (P) and negative (N) versions of this SST anomaly pattern. The atmospheric response is similar to the classical NAO pattern, although there are differences such as the centre of the 'Azores high' being about 5° too far north. A pressure difference of 6 hPa between the Azores and Iceland (Fig. 3b) is comparable with the standard deviation of the observed DJF NAO index. Surface temperature changes (not shown) over northern Europe, the eastern United States, Greenland and the Middle East are all consistent with the expected local thermal advection by the NAO wind field²⁰.

The same MSLP pattern is obtained when DJF MSLP from the ensemble of six simulations is regressed onto the time series of the SST anomaly pattern. The correlation between the SST pattern time series and the unfiltered observed and ensemble average NAO is 0.53 and 0.47, respectively (statistically significant at the 99.9% level). Multiple regression of the NAO onto time series of the SST anomaly centres GL, US and ST (Fig. 3a) suggests that GL and US are particularly important for forcing the NAO.

We now investigate the mechanism by which this SST signal might affect the atmosphere. Figure 3c shows the difference in evaporation from the ocean surface between model simulations P and N. The flux of latent heat by evaporation represents an important heat flux at the surface^{7,8,17} and dominates other surface heat flux changes in P minus N. There is strong spatial coherence between Figs 3a and c, with increased evaporation in regions of positive SST anomalies and decreased evaporation in regions of negative SST anomalies. In middle latitudes, there is a secondary effect of a southward shift of the evaporation centre relative to the GL SST anomaly; this may be due to the increased sensitivity of evaporation to the warmer SSTs in the south of this region and also to feedbacks with the eastward wind. Figure 3d shows the difference in precipitation in terms of changes in the associated latent heat release and therefore in column-mean heating in the atmosphere. Comparing Fig. 3c and d, there is again strong spatial coherence with about 60% of the SST-forced changes in evaporation being 'rained-out' locally. Convective precipitation changes dominate this signal, although large-scale precipitation (slope convection) also makes an appreciable contribution in the middle latitudes.

As expected, changes in the atmospheric thermal structure, for instance in the thickness of the 200–925 hPa layer, are in quadrature with the changes in latent heating, with thermal anomalies centred immediately downstream of heating anomalies (Fig. 3d). This can be seen directly in the 200-hPa geopotential height anomalies, Fig. 3e, which are the dominant contributor to the thickness anomalies. Figure 3b and e indicate a near-equivalent barotropic vertical structure^{2,4,28} with a deep 'low' near Iceland and the band of high pressure near 45°N extending throughout the troposphere. This height field strongly resembles the 'NAO' modes mentioned by other authors⁴. In the 'Azores high' region, although there are

undoubtedly feedbacks within the atmosphere, the indication is that local SSTs play a role in forcing the observed cooling, Fig. 3d (with a maximum at about 790 hPa), and hence are also important for the low-level anticyclonic circulation centred around 30°W, 45°N (Fig. 3b).

Hence we have demonstrated that North Atlantic SST anomalies can lead to local changes in surface evaporation, precipitation and atmospheric heating that tend to re-enforce the thermal and geopotential structure of the NAO⁴. This forcing of the NAO is distinct from the natural tendency of the NAO to damp diabatically⁴. We also find a northward shift and 13% intensification into northern Europe of the stormtrack, Fig. 3f. Such a shift has been associated with the pressure field of positive phases of the NAO³,

although interactions between synoptic variations and the large-scale flow may complicate a simple interpretation⁴.

We now briefly examine the implied changes in heat fluxes in the oceanic mixed layer between simulations P and N. Evaporation (Fig. 3c), which represents a cooling of the ocean, generally acts to damp the P minus N SST anomalies (Fig. 3a). The same result emerges from a regression analysis of the ensemble of simulations. Hence a positive feedback on SSTs involving evaporation, considered by previous authors^{14,15}, does not occur in this model. Ekman transport is another important term in the oceanic mixed-layer heat budget^{7,17}. In the steady state, an anomalous eastward windstress in the Northern Hemisphere, for example, must be balanced by a Coriolis term, and hence by cold water advection from the north.

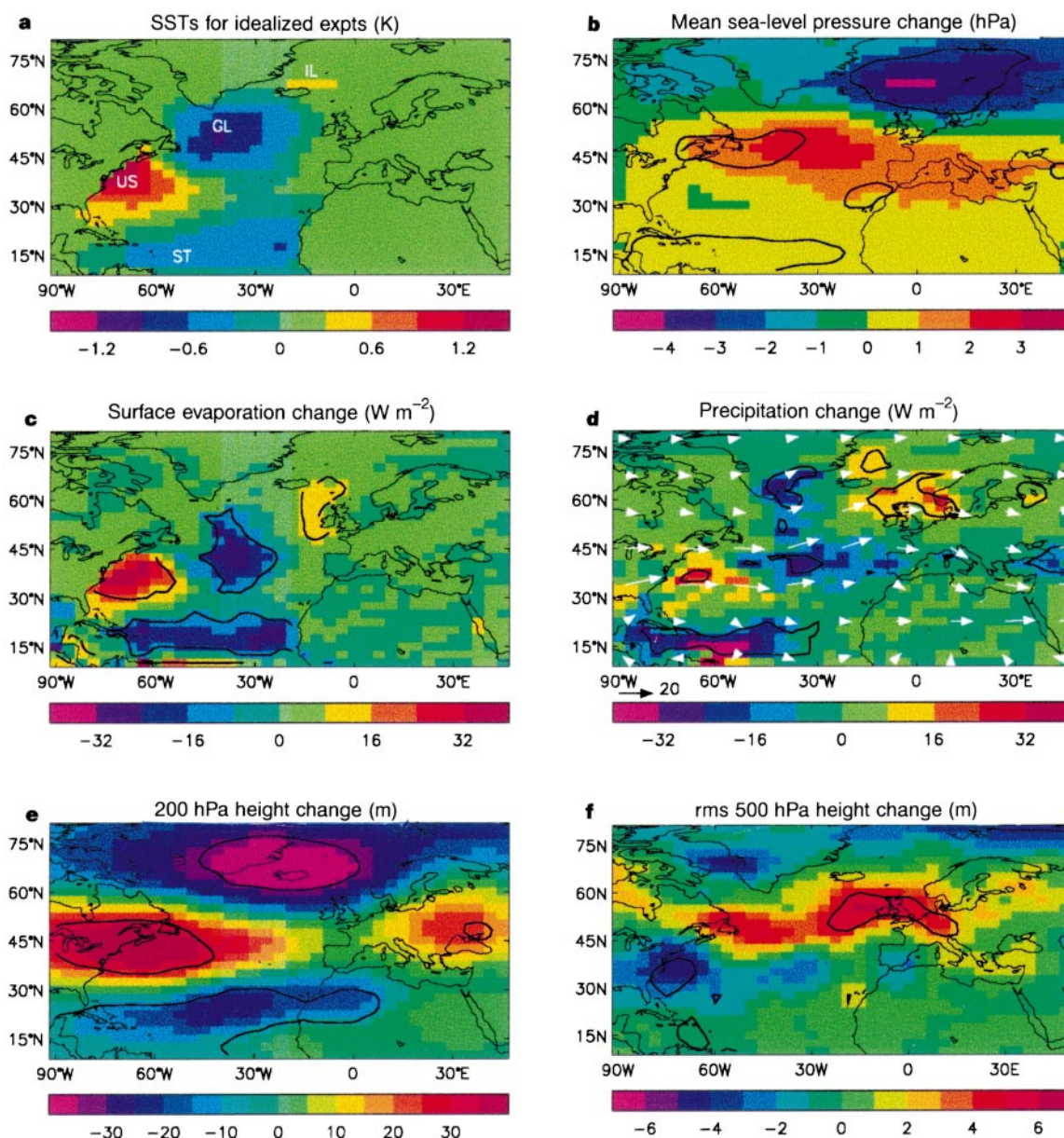


Figure 3 Model inputs and simulated responses. **a**, The pattern of SST anomalies used to force two 20-year model simulations, taken from a joint canonical correlation analysis on SSTs and MSLP¹⁶. The values shown correspond to the average of the monthly-varying SST anomalies used in the December to February season. For each month of the year, the scaling is calculated as two standard deviations of the time series obtained by projecting observed SSTs for that month onto the pattern. SST anomalies are added to (simulation P) or subtracted from (simulation N) a climatological SST field for 1961–90. **b**, December to February

MSLP difference, simulation P minus simulation N. **c**, As **b**, but for surface evaporation in terms of latent heat flux (W m^{-2}). **d**, As **b** but for total precipitation in terms of condensational latent heat release (W m^{-2}). Superimposed are mean 500-hPa wind vectors calculated as the average of simulations P and N. The reference vector is 20 m s^{-1} . **e**, As **b** but for 200-hPa geopotential height. **f**, As **b** but for the 20-season mean of the r.m.s. of 2.5–6 day bandpass-filtered²⁷ daily 500-hPa height. In **b–f**, the areas within the black contours exceed the 95% confidence level of non-zero difference using a 2-tailed *t*-test.

We find that the Ekman transport provides a positive feedback to the P minus N SST anomaly over almost the entire North Atlantic and for most of the year. North of about 45°N, this term often provides a positive feedback that is stronger than any negative feedback from evaporation. Elsewhere, evaporation tends to be stronger.

We have provided evidence that simulations of the NAO in the HadAM2b model are influenced by a pattern of ocean surface temperatures which is similar to one of the most important naturally occurring patterns of North Atlantic SST variability⁹. Observational studies have shown that there may be significant predictability of mixed-layer oceanic temperatures in the North Atlantic at multiannual timescales⁵ and these have been linked in observational studies to the NAO²⁹. This suggests the possibility of multiannual, as well as seasonal, predictions of European winter climate features such as surface temperature, precipitation and storminess. Our use of an atmospheric model forced with imposed SSTs was essential in establishing the chain of cause and effect from SST anomalies to NAO. Previous studies have suggested that extra-tropical ocean surface heat fluxes may be over-estimated in atmospheric model simulations forced with prescribed SSTs²⁸. Our results, however, imply that the extra-tropical (as well as tropical) evaporative heat flux, and the SSTs beneath them, may be central to the skill of our simulation of the NAO. Work is now required to resolve these differences and to understand the interannual and longer-timescale predictability of the SST pattern in Fig. 3a. This will involve using a fully coupled model^{28,30} where synoptic and longer-timescale forcings of the ocean by the atmosphere are included. □

Received 24 August 1998; accepted 29 January 1999.

- Kutzbach, J. E. Large-scale features of monthly mean Northern Hemisphere anomaly maps of sea-level pressure. *Mon. Weath. Rev.* **98**, 708–716 (1970).
- Wallace, J. M. & Gutzler, D. S. Teleconnections in the geopotential height field during Northern Hemisphere winter. *Mon. Weath. Rev.* **109**, 784–812 (1981).
- Rogers, J. C. Patterns of low-frequency monthly sea level pressure variability (1899–1986) and associated wave cyclone frequencies. *Q. J. R. Meteorol. Soc.* **122**, 1385–1414 (1996).
- Ting, M. & Lau, N.-C. A diagnostic and modelling study of the monthly mean wintertime anomalies appearing in a 100-year GCM experiment. *J. Atmos. Sci.* **50**, 2845–2867 (1993).
- Sutton, R. T. & Allen, M. R. Decadal predictability of North Atlantic sea surface temperature and climate. *Nature* **388**, 563–567 (1997).
- Bjerknes, J. Atlantic air-sea interaction. *Adv. Geophys.* **10**, 1–82 (1964).
- Daly, A. W. The response of North Atlantic sea surface temperature to atmospheric forcing processes. *Q. J. R. Meteorol. Soc.* **104**, 363–382 (1978).
- Cayan, D. R. Latent and sensible heat flux anomalies over the northern oceans: The connection to monthly atmospheric circulation. *J. Clim.* **5**, 354–369 (1992).
- Deser, C. & Blackmon, M. L. Surface climate variations over the North Atlantic ocean during winter: 1900–1989. *J. Clim.* **6**, 1743–1753 (1993).
- Namias, J. Seasonal persistence and recurrence of European blocking during 1958–1960. *Tellus* **16**, 394–407 (1964).
- Ratcliffe, R. A. S. & Murray, R. New lag associations between North Atlantic sea temperatures and European pressure, applied to long-range weather forecasting. *Q. J. R. Meteorol. Soc.* **96**, 226–246 (1970).
- Palmer, T. N. & Sun, Z. A modelling and observational study of the relationship between sea surface temperature in the northwest Atlantic and the atmospheric general circulation. *Q. J. R. Meteorol. Soc.* **111**, 947–975 (1985).
- Davies, J. R., Rowell, D. P. & Folland, C. K. North Atlantic and European seasonal predictability using an ensemble of multi-decadal AGCM simulations. *Int. J. Climatol.* **17**, 1263–1284 (1997).
- Latif, M. & Barnett, T. P. Causes of decadal variability over the North Pacific and North America. *Science* **266**, 634–637 (1994).
- Grötzner, A. M., Latif, M. & Barnett, T. P. A decadal climate cycle in the North Atlantic ocean as simulated by the ECHO coupled GCM. *J. Clim.* **11**, 831–847 (1998).
- Frankignoul, C. & Hasselmann, K. Stochastic climate models, Part II: Application to sea-surface temperature anomalies and thermocline variability. *Tellus* **29**, 289–305 (1977).
- Frankignoul, C. Sea surface temperature anomalies, planetary waves and air-sea feed-back in the middle latitudes. *Rev. Geophys.* **23**, 357–390 (1985).
- Hall, C. D., Stratton, R. A. & Gallani, M. L. *Climate Simulations with the Unified Model: AMIP Runs* (Clim. Res. Tech. Note 61, UK Meteorol. Office, Bracknell, 1995).
- Lamb, P. J. On the mixed-layer climatology of the north and tropical Atlantic. *Tellus A* **36**, 292–305 (1984).
- Walker, G. T. & Bliss, E. W. World Weather V. *Mem. R. Meteorol. Soc.* **4**, 53–84 (1932).
- Van Loon, H. & Rogers, J. C. The seesaw in winter temperatures between Greenland and northern Europe. Part I: General description. *Mon. Weath. Rev.* **106**, 296–310 (1978).
- Hurrell, J. W. Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. *Geophys. Res. Lett.* **23**, 665–668 (1996).
- Hulme, M. A 1951–80 global land precipitation climatology for the evaluation of general circulation models. *Clim. Dyn.* **7**, 57–72 (1992).
- Van den Berg, W. D. The role of various weather parameters and the use of worst-case forecasts in prediction of gas sales. *Meteorol. Appl.* **1**, 33–37 (1994).
- Kushnir, Y. Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Clim.* **7**, 141–157 (1994).
- Xie, S. P. A pan-Atlantic decadal climate oscillation. *Geophys. Res. Lett.* **25**, 2185–2188 (1998).
- Blackmon, M. L. & Lau, N.-C. Regional characteristics of the Northern Hemisphere wintertime

circulation: A comparison of the simulation of a GFDL general circulation model with observations. *J. Atmos. Sci.* **37**, 497–514 (1980).

- Saravanan, R. Atmospheric low-frequency variability and its relationship to midlatitude SST variability: Studies using the NCAR climate system model. *J. Clim.* **11**, 1386–1404 (1998).
- Molinari, R. L., Mayer, D. A., Festa, J. F. & Bezdek, H. F. Multiyear variability in the near-surface temperature structure of the midlatitude western North Atlantic ocean. *J. Geophys. Res.* **102**, 3267–3278 (1997).
- Barsugli, J. J. & Battisti, D. S. The basic effects of atmosphere–ocean thermal coupling on midlatitude variability. *J. Atmos. Sci.* **55**, 477–493 (1998).

Acknowledgements. We thank C. Gordon, T. Basnett, C. Cooper, B. Horton, W. Ingram and D. Sexton for their help and comments. This work was supported by the Public Meteorological Service; computer time was provided by the Department of the Environment, Transport and the Regions.

Correspondence and requests for materials should be addressed to M.J.R. (e-mail: mjrodwell@meto.gov.uk).

Distortion of isochronous layers in ice revealed by ground-penetrating radar

David G. Vaughan*, Hugh F. J. Corr*, Christopher S. M. Doake* & Ed. D. Waddington†

* British Antarctic Survey, Madingley Road, Cambridge CB3 0ET, UK

† Geophysics Program, University of Washington, Box 351650, Seattle, USA

In addition to measuring ice-sheet thickness, ground-penetrating radar can be used to delineate reflections in ice sheets^{1–3}. These reflections are generally accepted to result from layers of isochronous deposition of snow and can reveal much about the dynamics of the ice flow. Here we present ground-penetrating radar data from Fletcher Promontory, Antarctica, which show arches and troughs in isochronous ice layers to a depth of 100 m. We demonstrate that the origin of these features can be determined by their growth with depth, and many features result from local anomalies in accumulation rate which can be correlated with the ice surface slope. One arch appears to be the result of a local anomaly in vertical strain-rate. Its proximity to the ice divide, width, and growth with depth, indicate that this arch is one of a class of features postulated by Raymond⁴ but not previously shown to exist in the field. Such a feature is an indication of the nonlinear rheology of ice and requires that palaeoclimate records from ice cores extracted from the vicinity of ice divides underlain by similar features should be specifically corrected for such effects on ice-flow.

Fletcher Promontory, Antarctica, is a 500-m thick ice cap, atop a shear-sided, inclined mesa between Rutford Ice Stream and Carlson Inlet (Fig. 1). The mean annual air temperature is about –20 °C and ice flow is only a few metres per year (ref. 5), indicating that the ice is frozen to its bed. Snow accumulation close to the crest was measured over one year as $(320 \pm 60) \text{ kg m}^{-2} \text{ yr}^{-1}$.

Figure 2 shows a 15-km-long ground-penetrating radar (GPR) section collected over Fletcher Promontory. The section shows many internal horizons in the uppermost 100 m, which are thought to result from density variations related to depositional processes^{6,7}; but as the wavelengths used are greater than the thickness of the snow layer from any one precipitation event, each internal horizon corresponds to interference from several layers⁸. Indeed, some layers were resolved into multiple horizons by using 200-MHz GPR data not presented here.

The internal horizons shown in Fig. 2b undulate to give arches and troughs which are unrelated to bed topography. Figure 3 shows the dip of predicted particle paths across Fletcher Promontory, together with the dip of lines drawn down through arches and troughs. The agreement suggests that the arches and troughs result from burial processes, not from the flow of ice over topography.

Defining the x-axis as horizontal and the z-axis as vertically downwards, the burial rate of ice from the surface ($x = 0$, $z = 0$)

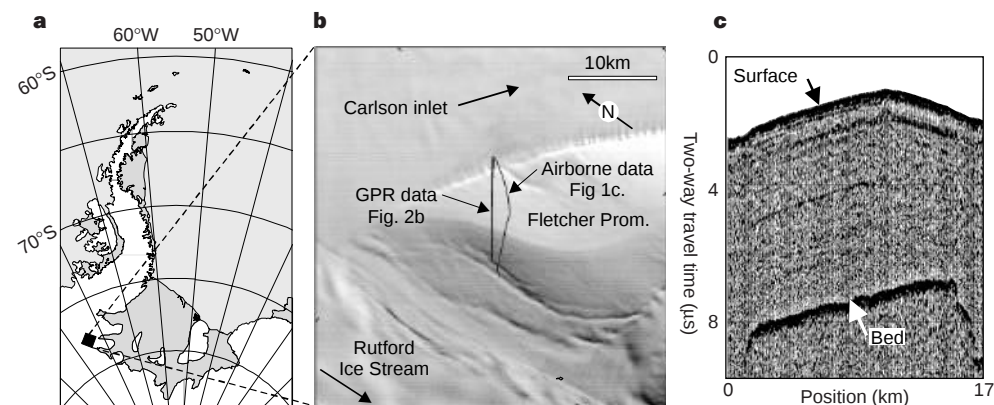


Figure 1 Orientation of Fletcher Promontory. **a**, Location map showing area covered by subscene. **b**, Landsat MSS subscene (path/row 225/117, 3 February 1974). The location of the GPR section (15 km long) and flow directions of Rutford Ice Stream (flowing at $>300 \text{ myr}^{-1}$) and Carlson Inlet (at present almost stagnant²⁶) are shown. **c**, Airborne radar data showing surface and bed topography over the track indicated in **b**.

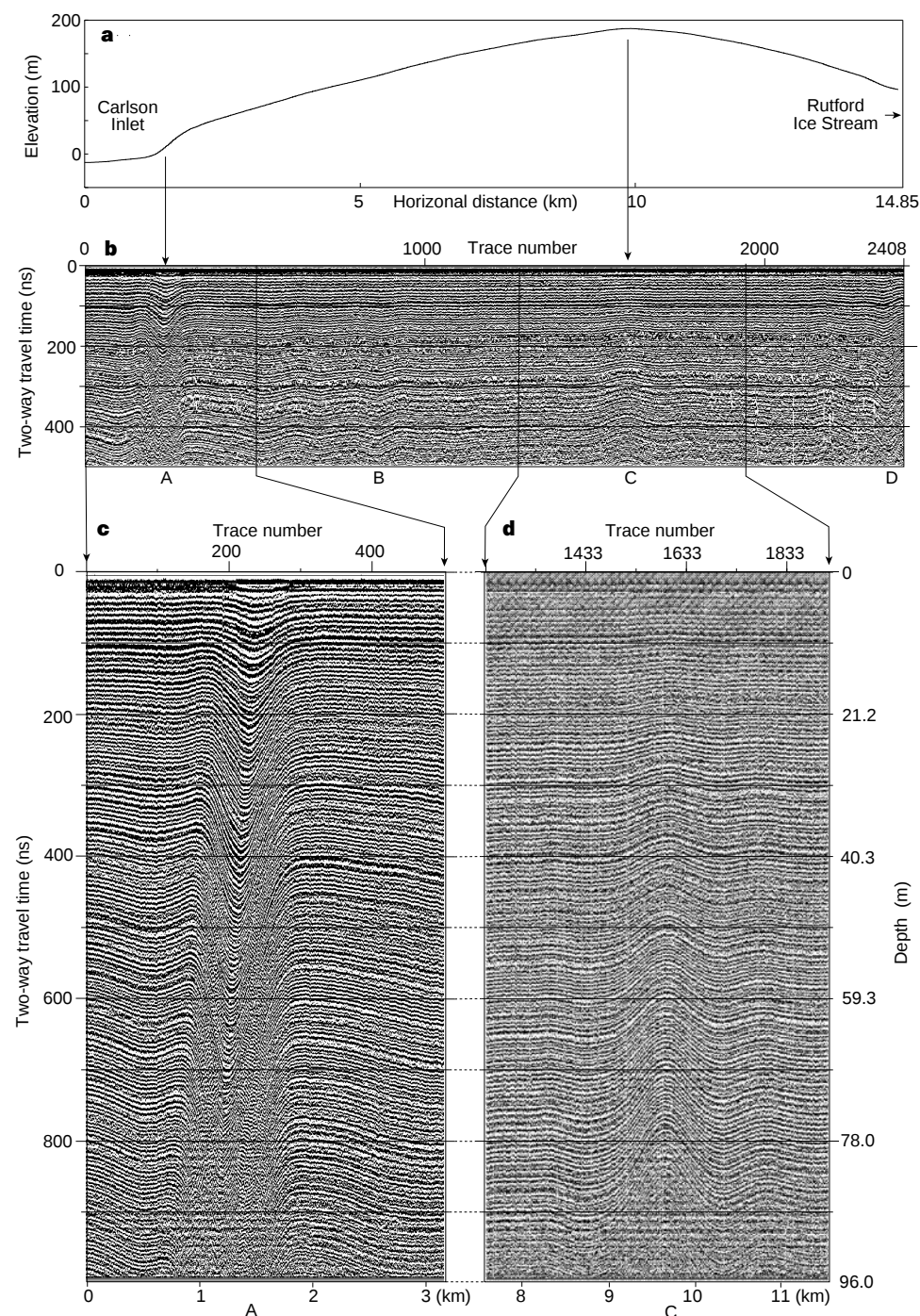


Figure 2 Results of the GPR and GPS, Global Positioning System, investigations on Fletcher Promontory. **a**, Surface topography along the track indicated in Fig. 1, measured using carrier-phase GPS. **b**, 100-MHz GPR section acquired using separate transmitting and receiving antennas giving a pulse length of $\sim 2.5 \text{ m}$ in ice, mounted with a 1-m offset and towed at $\sim 10 \text{ km h}^{-1}$. GPR samples (traces) were collected at equal time intervals along the track, not at equal distance intervals. The locations of features at sites A–D, which yield the burial curves shown in Fig. 3, are shown. **c**, Expanded view of the arch at site A, which coincides with the area of high surface slope. **d**, Expanded view of the arch at site C, which coincides with the topographic crest. In **c** and **d**, common midpoint measurements were used to determine a velocity/depth relation, used to convert two-way travel times to depth. Depth was converted to ice-equivalent depth, using the relation given by Paterson, with surface density of 350 kg m^{-3} , and a depth to firn transition of 70 m (ref. 27, page 14).

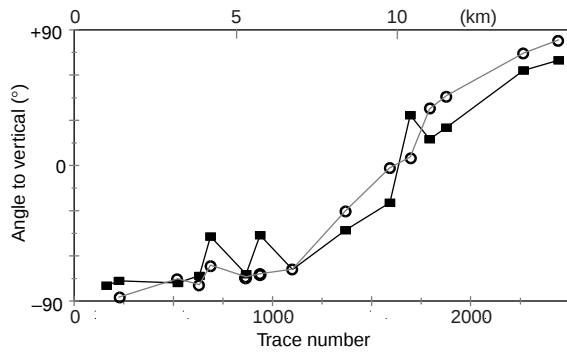


Figure 3 Burial trajectories of GPR features. The angle to the vertical of loci joining crests of arches and bases of troughs in the isochronous layers (filled squares) can be compared with the angle to the vertical of predicted particle paths (open circles). Particle paths were predicted assuming that ice velocity perpendicular to the surface is equal to the accumulation rate, and that ice velocity parallel to the surface can be obtained from the surface slope (ref. 27, page 251) using the following flow-law parameters; $A = 1.7 \times 10^{-16} \text{ kPa}^{-3} \text{ s}^{-1}$ and $n = 3$, and ice density $= 917 \text{ kg m}^{-3}$. The equation used to calculate velocity is invalid near trace no. 200 where ice velocity is perpendicular to the profile.

to ice-equivalent depth, z , is given by

$$\frac{dz}{dt} = \dot{a}_0 + \dot{a}'_0 t + \dot{\epsilon}_0 z \quad (1)$$

where $\dot{a}_0$ is the accumulation rate, $\dot{a}'_0$ is the accumulation-rate gradient assumed constant in the region traversed by the ice particles, $\dot{\epsilon}_0$ is the vertical strain-rate (negative indicating thinning layers), and u_0 is the horizontal velocity (see Supplementary Information). If we assume that u_0 is constant within the surface layers, and that $\dot{\epsilon}_0$ varies only in x , equation (1) has a solution

$$z(x_0, t) = \frac{\dot{a}_0}{\dot{\epsilon}_0} [e^{\dot{\epsilon}_0 t} - 1] + \frac{\dot{a}'_0 u_0}{\dot{\epsilon}_0^2} [e^{\dot{\epsilon}_0 t} - 1 - \dot{\epsilon}_0 t] \quad (2)$$

which satisfies the surface boundary condition, $z = 0$ where $t = 0$. On adjacent particle paths where an isochronous layer is buried to different depths forming an arch or trough, its size, δz , can be given as

$$\delta z = z(x_0, t) - z(x_1, t) \quad (3)$$

where the subscripts denote horizontal positions at the point of burial. If the different burial depths result from different accumulation rates ($\dot{a}_0$ and $\dot{a}_1$) while vertical strain-rate ($\dot{\epsilon}_0$) is constant, then:

$$\delta z \approx \frac{(\dot{a}_0 - \dot{a}_1)}{\dot{a}_0} z + \left[(\dot{a}_0 - \dot{a}_1) \frac{\dot{\epsilon}_0}{\dot{a}_0^2} + \frac{\dot{a}'_1 u_1}{\dot{a}_0^2} \right] \frac{z^2}{2} \quad (4)$$

Alternatively, if the different burial depths result from different vertical strain-rates ($\dot{\epsilon}_0$ and $\dot{\epsilon}_1$) while the accumulation rate ($\dot{a}_0$) is constant, then:

$$\delta z = \frac{1}{2\dot{a}_0} (\dot{\epsilon}_0 - \dot{\epsilon}_1) z^2 \dots + \text{higher terms} \quad (5)$$

With a strain rate of 10^{-3} yr^{-1} , the higher terms which grow with depth will be around 5% at 50 m and 10% at 100 m.

Equations (4) and (5) together provide a test for the origin of particular features. Features whose amplitude initially increases linearly with depth then falls off are predicted by equation (4), whereas quadratic growth increasing with depth is predicted by equation (5). Figure 4 shows the test applied to features at sites A–D (the locations of these sites are shown in Fig. 2).

Features at sites A, B and D are typical of most of the section and follow equation (4), indicating they were formed by variations in accumulation rate. Figure 5 shows that this variation is correlated with the surface slope. Similar results have already been reported on

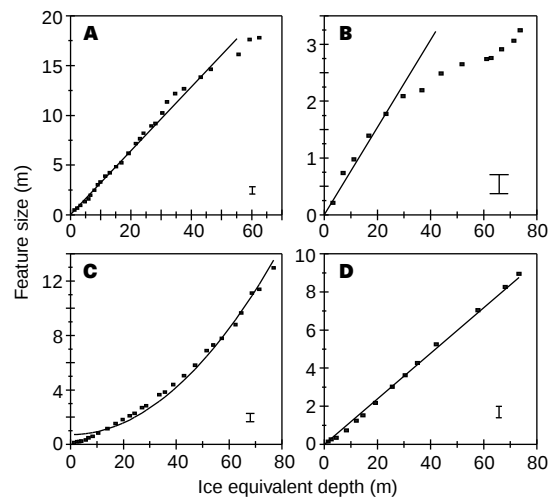


Figure 4 Amplitude of GPR features as a function of depth. Measurements from the GPR section of the size of arches and troughs at sites shown in Fig. 2 are shown against ice equivalent depth by filled squares. Features at sites A, B and D are fitted using equation (4) (solid lines) with accumulation anomalies of +33% at A, +8% at B and +4% at D. The feature at site C is fitted using equation (5) (solid line). In each panel, the 'error bars' at lower right indicate the uncertainty in determining the feature size (half the radar wavelength, $\sim 0.8 \text{ m}$).

other length-scales, and the processes likely to be responsible are deposition and aeolian redistribution^{9,10}.

This is significant, because we expect the particle velocity to be determined by conditions within the ice (such as temperature, and basal and surface slope). In the steady state, however, the particle velocity perpendicular to the surface must be matched by the accumulation rate ($\dot{a}$) which we now find depends on surface slope. Thus two independent conditions relate surface slope to accumulation rate; this implies that, on some length-scales, ice sheets may never achieve steady state but may support moving topographic waves⁹. We need to understand these waves if observations of non-steady surface elevation (for example, ref. 11) are to be taken as an indication of significant change.

The arch at site C (Fig. 2) lies beneath ($\pm 100 \text{ m}$) the topographic crest. Its amplitude increases quadratically with depth and is well fitted by equation (5), suggesting that it results from a vertical strain-rate anomaly of $(1.4 \pm 0.2) \times 10^{-3} \text{ yr}^{-1}$ —typical for ice creeping without crevassing at -20°C (ref. 12).

Raymond⁴ modelled conditions close to ice divides, predicting that arches of this type should indeed be present; in line with previous discussions we shall use the term 'Raymond bumps' to describe them. The formation of such bumps can be summarized thus. The basal driving stress is dependent on the surface slope and approaches zero near an ice divide. When frozen to the bed, these lower layers suffer unusually low deviatoric stress and so under a nonlinear rheology exhibit an unusually high viscosity resisting vertical deformation. Consequently, the surface layers undergo more rapid thinning than similar layers away from the divide, causing apparent up-arching of isochronous layers under the divide. A scaling analysis showed that the area affected should be approximately two ice thicknesses wide⁴. Raymond's predictions have since been refined by further modelling^{13,14}, and an inverse feature likely to occur at sliding/no-sliding transitions has also been discussed¹⁵.

Features that might be Raymond bumps have been reported on Berkner Island, Antarctica^{16,17} and on Siple Dome, Antarctica¹⁸, but these are at considerable depth, horizontally displaced from the present divide, and could also have occurred as a result of preferential wind-scouring at the ice divide¹⁹. Conversely, modelling²⁰ which suggested that Raymond bumps would be present at the

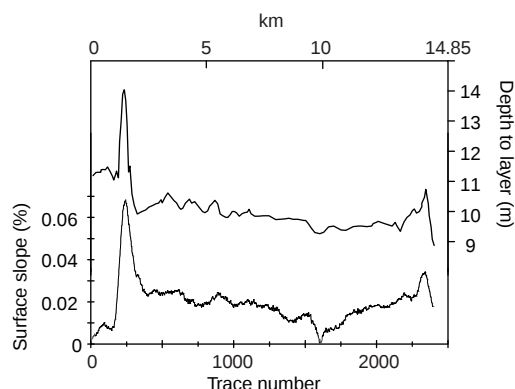


Figure 5 Comparison of accumulation rate with surface slope. The depth to a prominent internal horizon in Fig. 2b is taken as a proxy of accumulation rate. This depth is a good match to the magnitude of surface slope smoothed with a 120-m filter. Cross-correlation suggests that a 7% surface slope produces an accumulation rate ~30% higher than the regional average.

summit of Greenland was not supported by radar data^{21,22}. The arches presented here are, to our knowledge, the first to be reported in the near-surface layers, at the divide, and the first to be shown to result from anomalous strain rate. We note that the width of these arches ($1,570 \pm 50$ m) is approximately three times the ice thickness, which constitutes good agreement with Raymond's scaling analysis. The locus joining the crest of the arches tilts at -40° to the vertical, a probable indication of migration (0.5 m yr^{-1}) of the ice divide towards Rutford Ice Stream.

Although some benefits arise from collecting ice cores at ice divides, there are now at least two reasons to be wary of such sites; first, preferential wind-scouring which distorts isotopic ratios²³ and accumulation rates, and second, the possible presence of Raymond bumps. One precaution would be to ensure that flow models predicting depth–age relationships also reproduce the patterns in isochronous layers. Failure to account for Raymond bumps could incur substantial errors in accumulation-rate histories; for example, a hypothetical 100-m ice core from the crest of Fletcher Promontory would, unless corrected, yield erroneously low accumulation rates; an error of -2.6% at 10 m depth, -8% at 30 m and -26% at 100 m.

It is not yet clear over what length-scales and under what temperature conditions Raymond bumps form. But the presence of such bumps on Fletcher Promontory is a clear illustration of the nonlinear rheology of ice in a natural setting, which has often been questioned^{24,25}, and shows that special conditions can prevail beneath ice divides which should not be ignored when analysing even short ice cores. □

Received 16 April 1998; accepted 7 January 1999.

1. Robin, G. de Q. *The Climatic Record in Polar Ice Sheets* (Cambridge Univ. Press, 1983).
2. Clarke, T. S. & Bentley, C. R. High-resolution radar on Ice Stream B2, Antarctica: measurements of electromagnetic wave speed in firn and strain history from buried crevasses. *Ann. Glaciol.* **20**, 153–159 (1994).
3. Kohler, J., Moore, J., Kenet, M., Engeset, R. & Elvehøy, H. Using ground-penetrating radar to image previous year's summer surfaces for mass balance measurements. *Ann. Glaciol.* **24**, 255–260 (1997).
4. Raymond, C. F. Deformation in the vicinity of ice divides. *J. Glaciol.* **29**, 357–373 (1983).
5. Frolich, R. M. & Doake, C. S. M. SAR interferometry over Rutford Ice Stream and Carlson Inlet, Antarctica. *J. Glaciol.* **44**, 77–92 (1998).
6. Paren, J. G. & Robin, G. de Q. Internal reflections in polar ice sheets. *J. Glaciol.* **14**, 251–259 (1975).
7. Miners, W. D. *et al.* Forward modelling of the internal layers in radio echo sounding using electrical and density measurements from ice cores. *J. Phys. Chem. B* **101**, 6201–6204 (1997).
8. Harrison, C. H. Radio echo sounding of horizontal layers in ice. *J. Glaciol.* **12**, 383–397 (1973).
9. Black, H. P. & Budd, W. Accumulation in the region of Wilkes Land, Antarctica. *J. Glaciol.* **5**, 3–15 (1964).
10. Whillans, I. M. Effect of inversion winds on topographic detail and mass balance on inland ice sheets. *J. Glaciol.* **14**, 85–90 (1975).
11. Thomas, R. H., Stephenson, S. N., Bindschadler, R. A., Shabtaie, S. & Bentley, C. R. Thinning and grounding line retreat on the Ross Ice Shelf, Antarctica. *Ann. Glaciol.* **11**, 165–172 (1988).
12. Vaughan, D. G. Relating the occurrence of crevasses to surface strain rates. *J. Glaciol.* **39**, 255–267 (1993).
13. Dahl-Jensen, D. Two-dimensional thermo-mechanical modelling of flow and depth-age profiles near the ice divide in Central Greenland. *Ann. Glaciol.* **12**, 31–36 (1989).
14. Schott Hvidberg, C. Steady-state thermomechanical modelling of ice flow near the centre of large ice sheets with the finite-element technique. *Ann. Glaciol.* **23**, 116–123 (1996).

15. Weertman, J. Sliding-no sliding zone effect and age determination of ice cores. *Quat. Res.* **6**, 203–207 (1976).
16. Sandhager, H. Review of the Münster Airborne Radio-echo Sounding-data Set: Marine Ice beneath Filchner-Schelfeis; Bottom Reflectivity and Internal Structures of Berkner Island 111–114 (Filchner-Ronne Ice Shelf Programme Rep. No. 9, Alfred-Wegener Inst., Bremerhaven, 1996).
17. Steinhage, D. & Blindow, N. Results of Short Pulse Radio Echo Sounding on Thysenhöhe, Berkner Island, Antarctica 106–109 (Filchner-Ronne Ice Shelf Programme Rep. No. 10, Alfred-Wegener Inst., Bremerhaven, 1997).
18. Raymond, C. F. *et al.* Geometry and stratigraphy of Siple Dome, Antarctica. *Antarct. J. US* **30**, 91–93 (1995).
19. Nereson, N. A., Raymond, C. F., Waddington, E. D. & Jacobel, R. W. Recent migration of the Siple Dome ice divide, West Antarctica. *J. Glaciol.* **44**, 643–652 (1998).
20. Schott, C., Waddington, E. D. & Raymond, C. F. Predicted time-scales for GISP2 and GRIP boreholes at Summit, Greenland. *J. Glaciol.* **38**, 162–168 (1992).
21. Hempel, L. & Thyssen, F. Deep radio echo soundings in the vicinity of GRIP and GISP2 drill sites, Greenland. *Polarforschung* **62**, 11–16 (1992).
22. Jacobel, R. W. & Hodge, S. M. Radar internal layers from the Greenland summit. *Geophys. Res. Lett.* **22**, 587–590 (1995).
23. Fisher, D. A. *et al.* Effect of wind scouring on climatic records from ice-core oxygen-isotope profiles. *Nature* **301**, 205–209 (1983).
24. Doake, C. S. M. & Wolff, W. W. Flow law for ice in polar ice sheets. *Nature* **314**, 255–257 (1985).
25. Wolff, E. W. & Doake, C. S. M. Implications of the form of the flow law for a vertical velocity and age-depth profiles in polar ice. *J. Glaciol.* **32**, 366–370 (1986).
26. Frolich, R. M., Vaughan, D. G. & Doake, C. S. M. Flow of Rutford Ice Stream and comparison to Carlson Inlet, Antarctica. *Ann. Glaciol.* **12**, 51–56 (1989).
27. Paterson, W. S. B. *The Physics of Glaciers* 3rd edn (Elsevier Science, Oxford, 1994).

Supplementary information is available on Nature's World-Wide Web site and as paper copy from the London editorial office of Nature.

Acknowledgements. We thank C. Raymond, R. C. A. Hindmarsh, N. Nereson and colleagues at the British Antarctic Survey for their help and comments.

Correspondence and requests for materials should be addressed to D.G.V. (e-mail: d.vaughan@bas.ac.uk).

A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton

Ji Qiang[†], Luo Zhexi[‡] & Ji Shu-an^{*}

^{*} National Geological Museum of China, Beijing 100034, China

[†] China University of Geosciences, Beijing 100083, China

[‡] Section of Vertebrate Paleontology, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213-4080, USA

Here we describe a new triconodont mammal from the Late Jurassic/Early Cretaceous period of Liaoning, China. This new mammal is represented by the best-preserved skeleton known so far for triconodonts which form one of the earliest Mesozoic mammalian groups with high diversity. The postcranial skeleton of this new triconodont shows a mosaic of characters, including a primitive pelvic girdle and hindlimb but a very derived pectoral girdle that is closely comparable to those of derived therians. Given the basal position of this taxon in mammalian phylogeny, its derived pectoral girdle indicates that homoplasies (similarities resulting from independent evolution among unrelated lineages) are as common in the postcranial skeleton as they are in the skull and dentition in the evolution of Mesozoic mammals. Limb structures of the new triconodont indicate that it was probably a ground-dwelling animal.

Class Mammalia

Infraclass Triconodonta (McKenna and Bell 1997)

Order Eutriconodonta (Kermack *et al.* 1973)

Family Incertae sedis

Jeholodens jenkinsi gen. et sp. nov.

Etymology. *Jehol*: an ancient geographic name for the western part of the Liaoning Province, China; the namesake of the Jehol fauna in the Yixian Formation that yielded the holotype; *odens* (Latin): tooth; *jenkinsi* (Latin): in honour of F. A. Jenkins Jr for his pioneer studies of the evolutionary morphology of the mammalian postcranial skeleton.

Holotype. GMV 2139 a, b, a nearly complete skeleton consisting of a partial skull and all of the postcranial skeleton preserved as two



jugal; la, lacrimal; L1–L7, lumbar vertebrae 1–7; ma, metacromion (on the spine of the scapula); mf, mandibular foramen; mg, meckelian groove; m1–4, molars 1–4; mp1–5, metacarpals 1–5; mt1–mt5, metatarsals 1–5; mx, maxillary; n, nasal; pb, pubic; pc, pars cochlearis of petrosal; pcd, coronoid process of dentary; pe, petrosal, preserved in the dorsal/odontocephalic view; pf, pterygoid fossa; ph1–5; phalanges 1–5; pmx, premaxillary; ps, pterygoid shelf on the ventral border of the mandible; p1, p2, premolars 1, 2; ra, radius; r1–r15, thoracic ribs 1–15; sc, scapula; sn, scapular notch; so, supraoccipital; sp, spine of scapula; sq, squamosal; ss, supraspinous fossa of scapula; stb2–6, sternbrae 2–6 (the interclavicle, the sternal manubrium and sternbra 1 are preserved and exposed by preparation, but are not shown in **a**); sym, mandibula symphysis; s1, s2, sacral vertebrae 1, 2; ti, tibia; tm, teres major fossa (on scapula); t1–t15, thoracic vertebrae 1–15; ug1–5, ungual phalanges 1–5; ul, ulna; x, xiphoid process of sternum.

counterparts (Fig. 1a; a reconstruction of the specimen is shown in Fig. 1b).

Locality and horizon. The Sihetun site (at roughly 41° 40' 12" N, 120° 47' 36" E), about 32 km east of Chaoyang City, Liaoning Province, China. The holotype is from the lacustrine shales that are intercalated with neutrobasic volcanic beds in the Yixian Formation.

Geological age and fauna. Correlation of the Yixian Formation is equivocal. It has been suggested to be Late Jurassic¹, near the Jurassic–Cretaceous boundary^{2–4}, or Early Cretaceous⁵. The associated fauna includes the theropods *Sinosauropteryx*⁴, *Protarchaeopteryx*⁵, *Caudipteryx*⁵, the birds *Confusiusornis* and *Liaoningornis*⁵, the pterosaur *Eosipterus*, the mammal *Zhangheotherium*³, and diverse fossil fishes, invertebrates and plants.

Diagnosis. Dental formula 4.1.2.3/4.1.2.4 (incisors, canine, premolars, molars); linguobuccally compressed molars with three main cusps in a straight alignment (Fig. 2a, b); uniquely derived among triconodonts in possessing spoon-shaped incisors (Fig. 2c). *Jeholodens jenkinsi* differs from morganucodontids in lacking the molar-interlocking mechanism (by cingular cuspule e and cusp b) found in morganucodontids^{6–8}, in having weak labial cingula of the upper molars (Fig. 2a), in lacking lower cingular cuspules e, f and g (kuhnecone), and in lacking the angular process and the postdentary trough on the mandible (Fig. 2c). In *J. jenkinsi*, lower molar cusp a occludes into the valley-groove between cusps A and B of the opposite upper molar, thus differing from amphilestids and gobiconodontids^{9,10}, in which lower cusp a occludes into the embrasure anterior to upper cusp B. *J. jenkinsi* also lacks cingular cuspules e and f in the lower molars of the latter groups. The lower molars are interlocked, with a crescent-shaped distal cusp d of the preceding molar fitted into the concave mesial margin of cusp b of the succeeding molar, which is diagnostic of the Triconodontidae^{11–14}. The main cusp, a, of the new taxon is much higher than cusps b and c, a primitive character that is absent in triconodontids^{11–15}. *J. jenkinsi* differs from most triconodontids, other than *Alticonodon*¹¹ and *Ichthyconodon*¹⁴, in lacking the continuous lingual cingulum on the lower molars; it differs from *Alticonodon* and *Ichthyconodon* in having a shelf-like cusp d.

Lower incisor *i*₂ is about to be replaced by an erupting tooth (labelled *i*_{2r} in Fig. 2c) and *m*₄ is erupting in the holotype of

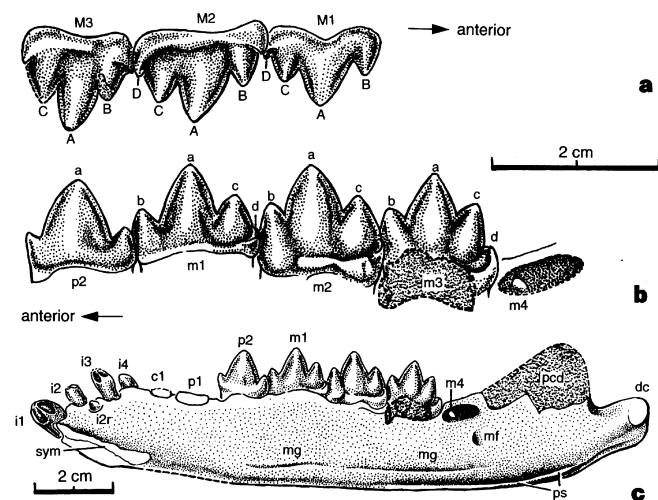


Figure 2 Dentition and mandible of *J. jenkinsi*. **a**, Upper molars M1–3 (right side labial view). **b**, Lower postcanines p₂–m₄ (right side, lingual view; m₄ is in the process of eruption). **c**, Mandible (right side, lingual view, corrected for slight distortion in the original specimen). Upper molars are labelled in the figure as M1–M3. Lower teeth are labelled as m1–m4; p1, p2, i1–i4 and i2r. We follow ref. 7 for the designation of cusps A–D on the upper molars and cusps a–d on the lowers. For other abbreviations see Fig. 1 legend. An analysis of mandibular and dental characters is found in Supplementary Information.

Jeholodens jenkinsi. This dental replacement indicates that the individual probably had not reached a full adult stage. We interpret the cusp pattern and the occlusion of laterally compressed molars in *J. jenkinsi* to be indicative of an insectivorous diet.

A noteworthy skeletal feature of this new taxon is a highly derived scapula (Figs 1a, 3a), confirming an earlier observation of a triconodontid from the Cretaceous of Montana¹². It has a robust, peg-like acromion and a low elevation on the spine that resembles the metacromion of *Didelphis* (Fig. 3f); the supraspinous fossa is fully developed. The dorsal part of the scapula has a prominent triangular area (Figs 1 and 3), similar to the large attachment area for the teres major muscle in monotremes¹⁶, a condition also present in the archaic therian *Zhangheotherium*³. The clavicle is curved. Its lateral end has a tapering point with a reduced contact to the acromion, whereas in living monotremes the two structures have a rigid and broad articulation. This suggests that *Jeholodens* had a mobile scapuloclavicular articulation. Medially, the clavicle has a limited overlap with the lateral process of the interclavicle, and lacks the rigid claviculointerclavicle articulation found in the tritylodont reptiles¹⁶ and monotremes¹⁷. We interpret this joint to have had at least some degree of mobility, like the conditions in multituberculates^{18–20} and in archaic³ and living²¹ therians.

The distal end of the humerus bears the radial and ulnar condyles on its anterior aspect, an incipient ulnar trochlea on its posterodorsal aspect, and reduced ectepicondyles and entepicondyles (on the counterpart slab; not shown), like the condition in *Zhangheotherium*³. These features differ from the primitive condition in cynodonts²², morganucodontids²³, multituberculates^{18–20}

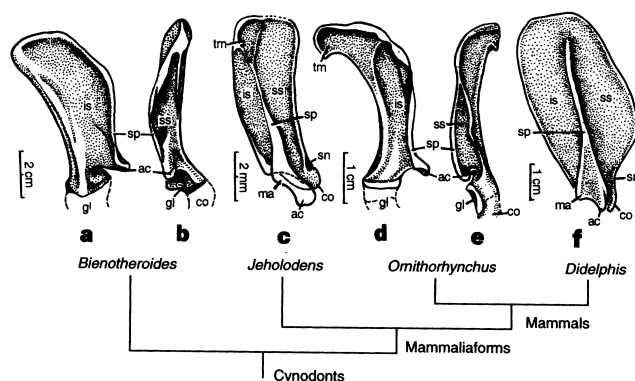


Figure 3 Mosaic evolution of the scapular characters in non-mammalian cynodonts and major mammalian clades. **a**, **b**, The tritylodontid *Bienotheroides* (lateral and anterior views of right scapula, after ref. 16, representing the outgroup condition of cynodonts). **c**, *J. jenkinsi* (lateral view, right scapula, composite reconstruction from both left and right scapulae of holotype GMV 2139 a). **d**, **e**, The monotreme *Ornithorhynchus* (Carnegie Museum specimen CM 1478; lateral and anterior views of right scapula). **f**, The marsupial *Didelphis* (lateral view, right scapula). For abbreviations, see Fig. 1 legend. Given the phylogeny in Fig. 5 and those of other references^{3,27}, the following derived characters of the pectoral girdle and forelimb of *J. jenkinsi* and therian mammals (including *Zhangheotherium*) would be best interpreted as convergences: a full supraspinous fossa; a protruding and laterally positioned acromion; coracoid fused to scapula; a ventral and uniformly concave scapular glenoid; mobile claviculoscapular and clavicle-interclavicle joints; a greatly reduced interclavicle; reduced ectepicondyles and entepicondyles of humerus; an incipient trochlea of humerus for the ulna. Alternatively, if the features of *J. jenkinsi* and therians are regarded as primitive for mammaliaforms as a whole, then those similarities between *Ornithorhynchus* and the outgroup tritylodonts would have to be interpreted as atavistic reversals in *Ornithorhynchus* to those of cynodonts; such features include: weak supraspinous fossa on the anterior aspect of scapula; acromion on the anterior margin of scapula; a lateral and saddle-shaped glenoid; and rigid and broad articulations of the clavicle-interclavicle and the scapula-clavicle. In either scenario, homoplasies are prominent in features of the pectoral girdles and forelimbs^{8,19,22}. Clade ranks following ref. 27; see also the alternatives given in refs 8, 29.

and living monotremes, in which the humerus has no ulnar trochlea. The epiphyseal suture is not present in the long bones in the holotype of *J. jenkinsi*.

In contrast to the derived forelimb and pectoral girdle of *Jeholodens jenkinsi*, its pelvic girdle, hindlimb and pes share many plesiomorphic characters with morganucodontids²³, tritylodontids¹⁶ and other cynodonts²². The epipubis is present. The patellar groove on the distal femur is far less developed than in monotremes, multituberculates and therians. The calcaneum (Fig. 4b) has a very short tubercle, a broad anterolateral ('peroneal') shelf, and extensive fibular and tibial facets, all of which are identical to those of morganucodontids²³ and non-mammalian cynodonts²², but different from the distinct peroneal process in monotremes, multituberculates^{18,19} and therians²⁴ and from the elongate tubercle of multituberculates and therians. The astragalus has a broad and uniformly convex tibial facet, but a weakly developed neck and navicular facet. The astragalus and calcaneum contact each other in juxtaposition; this condition is similar to that in non-mammalian cynodonts²² and morganucodontids²³, but different from the partial overlap of these two ankle bones in multituberculates and the complete overlap of these bones in archaic²⁴ and living therians (Fig. 4). Metatarsal 5 is offset from the cuboid (Fig. 4), allowing a wide range of abduction of the lateral pedal digits, as in morganucodontids²³, monotremes and multituberculates¹⁹. The manual and pedal ungual phalanges are slightly curved dorsally and concave on both sides, similar to those of most ground-dwelling small mammals^{19,25}. A flexor tubercle on the midpoint in most of the ungual phalanges indicates a certain degree of ability to flex the claws. *J. jenkinsi* lacks the sesamoid ossicle(s) of the monotremes, multituberculates and therians. We interpret *J. jenkinsi* to have been a ground-dwelling small mammal with a plantigrade gait (Fig. 1b) and some capability for climbing on uneven substrates²³, but not an arboreal mammal.

The eutriconodonts first appeared in the Middle Jurassic¹⁵ and were quite diverse, with a worldwide distribution, in the Cretaceous^{9–14}. Although abundantly represented by teeth^{9–15} and some cranial^{12,15,26} and postcranial⁹ materials, no fully articulated skeleton was known, until now, for this diverse group. In traditional classifications of early mammals that were based primarily on dentition^{6,12,15}, eutriconodonts and morganucodontids belonged to the order Triconodonta. The traditional grouping of 'triconodonts' (morganucodontids + eutriconodonts) is considered to be a

grade, rather than a monophyletic group, according to more recent phylogenetic analyses^{26–28}. Studies of some dental and cranial characters indicate that eutriconodonts may also be a heterogeneous group^{10,28}.

It has been proposed that the postcranial features of cynodonts and early mammals were subjected to functional constraints of locomotion and therefore susceptible to homoplasies^{19,22,29}. Several cladistic studies of cynodonts and early mammals that incorporated postcranial characters^{27,30} indicated that dental characters are highly homoplastic, whereas the postcranial characters can be very informative for higher-level phylogeny. Because of the paucity of the relatively complete postcrania of the basal mammals, it is unknown whether (or to what degree) homoplasies exist among different postcranial skeletal parts (such as the forelimb versus the hindlimb) in early mammalian evolution.

The discovery of the first fully articulated triconodont skeleton offers an unprecedented opportunity for a more comprehensive assessment of the relationships of triconodonts, combining all evidence of the dentition, basicranium and postcranium, and will also allow us to elucidate the pattern of postcranial evolution in early mammals. Our phylogenetic analysis has placed *J. jenkinsi* in the basal part of the mammalian phylogenetic tree, outside the crown group of extant Mammalia (Fig. 5). Available evidence indicates that *J. jenkinsi* is a eutriconodont and far more derived than morganucodontids. Within eutriconodonts, the new taxon is more closely related to the triconodontid clade than it is to amphilestids and gobiconodontids (see Supplementary Information). These results are consistent with the ideas that triconodonts are paraphyletic^{26–28} and that the triconodont-like molar cusp and occlusal patterns are characters for a functional grade.

Jeholodens jenkinsi shows a mosaic of derived, therian-like characters for most parts of the pectoral girdle (Fig. 3) and the humerus, but very primitive characters for the vertebral column, pelvic girdle, hindlimb and pes (Fig. 4). Our phylogenetic analysis indicates that many apomorphies of the pectoral girdle and forelimb in *J. jenkinsi* are independently derived and convergent with those of therians. Such apomorphies include a fully developed supraspinous fossa, the acromion and metacromion on the scapular spine, the incipient ulnar trochlea and the reduced epicondyles of the humerus. Given the phylogeny shown in Fig. 5, the mobile clavicle and scapula in *J. jenkinsi* would represent convergences with those of the multituberculate–therian clade^{3,20,27}. Thus these pectoral

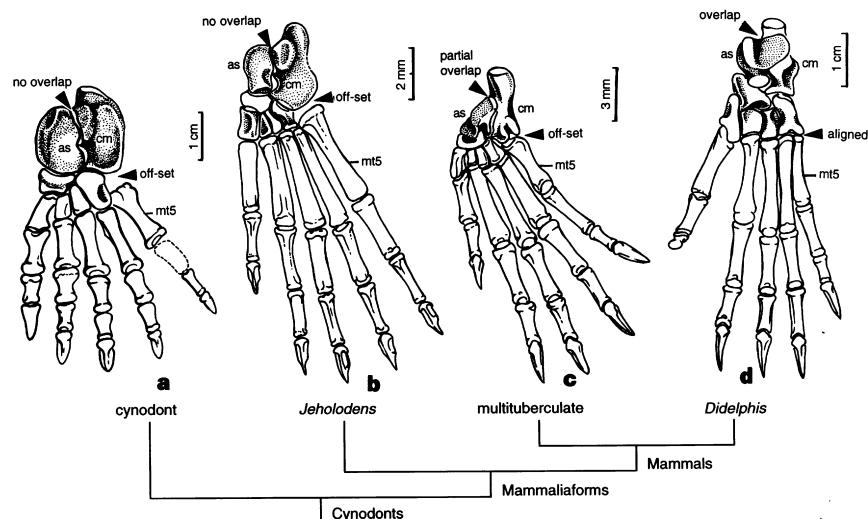


Figure 4 Pedes of mammals and outgroup cynodonts. **a**, A generalized cynodont condition (after refs 22, 28). **b**, *J. jenkinsi* (composite reconstruction of holotype GMV 2139 a,b). **c**, The multituberculate *Eucosmodon* (after refs 18, 19). **d**, The therian *Didelphis*. *J. jenkinsi* has little or no overlap of the astragalus and

calcaneum, a primitive feature of cynodonts, contrasting with the partial overlap of these characters in multituberculates and the complete overlap in therians. Clade ranks follow ref. 27; see also the alternatives given in refs 8, 29.

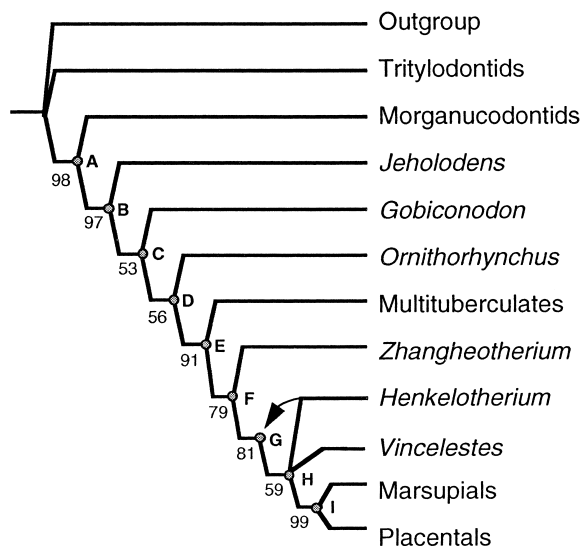


Figure 5 Evolutionary relationships of *J. jenkinsi*, gen. et sp. nov. The derived characters related to a mobile pectoral girdle occur separately on nodes B and E. The 'primitive' characters related to an immobile pectoral girdle occur in tritylodontids and the cynodont outgroup, and separately in *Ornithorhynchus* (node D). This indicates that there are many homoplasies of pectoral girdles and forelimbs (in contrast to the few or no convergences in the pelvic girdle and hindlimb), given the same tree topology of major clades of mammals. The arrow indicates an alternative placement of *Henkelotherium* at node G. For details of phylogenetic analysis see Methods and Supplementary Information.

characters, which allow a greater range of excursion of the shoulder joint in the locomotion of multituberculates^{3,19,20} and living therians²¹, have evolved at least twice among the Mesozoic mammals.

Alternatively, the mobile joints between the clavicle, interclavicle and scapula could be ancestral conditions shared by *J. jenkinsi* and the more derived mammals. If so, then the rigid clavicle–interclavicle articulation and relatively immobile scapula of monotremes (Fig. 3) would have to be regarded as atavistic reversals to the primitive conditions in the more distantly related non-mammalian cynodonts^{16,22}. For either evolutionary scenario, we must conclude that the pectoral girdles and forelimbs of early mammals underwent extensive convergent evolution, not only by comparison with the dental and cranial features, but also in relation to more conservative features of the pelvis and hindlimbs. □

Methods

Phylogeny of mammals (Fig. 5) is based on a strict consensus of two equally parsimonious trees (tree length = 210; consistency index = 0.638; retention index = 0.724) from PAUP analysis (3.1.1. Branch and Bound search) of 101 dental, cranial and postcranial characters that can be scored for the 12 major clades of mammals (see Supplementary Information). Most of the characters are preserved in the holotype of *Jeholodens jenkinsi*. The two most parsimonious trees differ only in the alternative placements of *Henkelotherium*, which either is at node G (represented by an arrow in Fig. 5) or switches positions with *Vincelestes*. These alternative placements of *Henkelotherium* do not alter the positions of any other clades, including *J. jenkinsi*. Numbers on branches represent the percentage of bootstrap values in 1,000 bootstrap replicas for a 50% majority bootstrap consensus tree that has identical topology to one of the two most parsimonious trees (that in which *Henkelotherium* is positioned at node G).

Received 26 August 1998; accepted 27 January 1999.

- Ji, Q. & Ji, S.-A. Discovery of the earliest bird fossils in China and the origin of birds. *Chinese Geol.* **1996**, 30–33 (1996).
- Hou, L., Martin, L. D., Zhou, Z. & Feduccia, A. Early adaptive radiation of birds: evidence from fossils from Northeastern China. *Science* **274**, 1164–1165 (1996).
- Hu, Y., Wang, Y., Luo, Z. & Li, C. A new symmetrodont mammal from China and its implications for mammalian evolution. *Nature* **390**, 137–142 (1997).

- Chen, P.-J., Dong, Z.-M. & Zhen, S.-N. An exceptionally well-preserved theropod dinosaur from the Yixian Formation of China. *Nature* **391**, 147–152 (1998).
- Ji, Q., Currie, P. J., Norell, M. A. & Ji, S.-A. Two feathered dinosaurs from northeastern China. *Nature* **393**, 753–761 (1998).
- Kermack, K. A., Mussett, F. & Rigney, H. W. The lower jaw of *Morganucodon*. *Zool. J. Linn. Soc.* **53**, 87–175 (1973).
- Crompton, A. W. The dentitions and relationships of the southern African Triassic mammals *Erythrotherium parringtoni* and *Megazostrodon rudnerae*. *Bull. Br. Mus. Nat. Hist.* **24**, 399–437 (1974).
- Luo, Z. in *In the Shadow of Dinosaurs—Early Mesozoic Tetrapods* (eds Fraser, N. C. & Sues, H.-D.) 980–128 (Cambridge Univ. Press, Cambridge, 1994).
- Jenkins, F. A. Jr & Schaff, C. R. The Early Cretaceous mammal *Gobiconodon* (Mammalia, Triconodontia) from the Cloverly Formation in Montana. *J. Vert. Paleontol.* **6**, 1–24 (1988).
- Kielan-Jaworowska, Z. & Dashzeveg, D. New Early Cretaceous amphirotid ('triconodont') mammals from Mongolia. *Acta Palaeont. Polonica* **43**, 413–438 (1998).
- Fox, R. C. Additions to the mammalian local fauna from the upper Milk River Formation (Upper Cretaceous), Alberta. *Can. J. Earth Sci.* **13**, 1105–1118 (1976).
- Jenkins, F. A. Jr & Crompton, A. W. in *Mesozoic Mammals: The First Two-thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 74–90 (Univ. Calif. Press, Berkeley, 1979).
- Cifelli, R. L., Wible, J. R. & Jenkins, F. A. Jr Triconodont mammals from the Cloverly Formation (Lower Cretaceous), Montana and Wyoming. *J. Vert. Paleontol.* **16**, 237–241 (1998).
- Sigogneau-Russell, D. Two possibly aquatic triconodont mammals from the Early Cretaceous of Morocco. *Acta Palaeont. Polonica* **40**, 149–162 (1995).
- Simpson, G. G. *A catalogue of the Mesozoic Mammalia in the Geological Department of the British Museum* (Oxford Univ. Press, London, 1928).
- Sun, A. & Li, Y. The postcranial skeleton of the late tritylodont *Bienotheroides*. *Vert. Palaeontol.* **23**, 136–151 (1985).
- Klima, M. Die Frühentwicklung des Schultergürtels und des Brustbeins bei den Monotremen (Mammalia: Prototheria). *Adv. Anat. Embryol. Cell Biol.* **47**, 1–80 (1973).
- Krause, D. W. & Jenkins, F. A. Jr The postcranial skeleton of North American multituberculates. *Bull. Mus. Comp. Zool.* **150**, 199–246 (1983).
- Kielan-Jaworowska, Z. & Gambaryan, P. P. Postcranial anatomy and habits of Asian multituberculate mammals. *Fossils Strata* **36**, 1–92 (1994).
- Sereno, P. & McKenna, M. C. Cretaceous multituberculate skeleton and the early evolution of the mammalian shoulder girdle. *Nature* **377**, 144–147 (1995).
- Jenkins, F. A. Jr & Weis, W. A. The functional anatomy of the shoulder in the Virginia opossum (*Didelphis virginiana*). *J. Zool.* **188**, 379–410 (1979).
- Jenkins, F. A. Jr The postcranial skeleton of African cynodonts. *Bull. Peabody Mus. Nat. Hist. Yale Univ.* **36**, 1–216 (1971).
- Jenkins, F. A. Jr & Parrington, F. R. Postcranial skeleton of the Triassic mammals *Eozostrodon*, *Megazostrodon*, and *Erythrotherium*. *Phil. Trans. R. Soc. Lond. B* **273**, 387–431 (1976).
- Rougier, G. W. *Vincelestes neuquenianus* Bonaparte (Mammalia, Theria), un primitivo mammifero del Cretacico Inferior de la Cuenca Neuquina. Thesis, Univ. Nacional de Buenos Aires (1993).
- McLeod, N. & Rose, K. D. Inferring locomotory behavior in Paleogene mammals via eigenshape analysis. *Am. J. Sci.* **293**, 300–355 (1993).
- Rougier, G. W., Wible, J. R. & Hopson, J. A. Basicranial anatomy of *Priacodon fruitaensis* (Triconodontidae, Mammalia) from the Late Jurassic of Colorado, and a reappraisal of mammalia-form interrelationships. *Am. Mus. Novit.* **3183**, 1–28 (1996).
- Rowe, T. Definition, diagnosis, and origin of Mammalia. *J. Vert. Paleontol.* **8**, 241–264 (1988).
- Kielan-Jaworowska, Z. Characters of multituberculates neglected in phylogenetic analyses of early mammals. *Lethaia* **29**, 249–255 (1997).
- Hopson, J. A. in *Major Features of Vertebrate Evolution* (eds Prothero, D. R. & Schoch, R. M.) 190–219 (Short Courses in Paleontol. No. 7, Paleontol. Soc. Knoxville, Tennessee, 1994).
- Kemp, T. S. The relationships of mammals. *Zool. J. Linn. Soc.* **77**, 353–384 (1983).

Supplementary information is available on Nature's World Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank K. C. Beard, R. L. Cifelli, W. A. Clemens, A. W. Crompton, M. R. Dawson, J. A. Hopson, F. A. Jenkins, Z. Kielan-Jaworowska, J. Meng, T. Rowe, D. Sigogneau-Russell, J. R. Wible and X.-c. Wu for discussions and for reviews of the manuscript; A. Henrici for preparation of the specimen; M. Klingler for preparing Fig. 1; and N. Wüetehle for assistance. This research was supported by funding from the Ministry of Geology and Mineral Resources of China and the National Natural Science Foundation of China (to J.Q.), and the US National Science Foundation, the National Geographic Society, and the M. Graham Netting Fund of the Carnegie Museum (to Z.L.).

Correspondence and requests for materials should be addressed to: Z. Luo (e-mail: luoz@clpgh.org).

Strong effects of weak interactions in ecological communities

E. L. Berlow

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA

The loss or removal of individual species can cause dramatic changes in communities^{1–5}. Experiments indicate that in many communities only a few species will have such strong effects, whereas most will have weak effects owing to small *per capita* effects and/or low abundance^{3,6–15,16}. But extinction of these 'weak' interactors could significantly alter natural communities because they play important stabilizing or 'noise-dampening' roles^{14,15,17–23}. I

demonstrate here that some 'weak' interactors may also be important by magnifying spatiotemporal variation in community structure. An analysis of published interaction strength data shows that the greatest variation in species effect occurred for the weakest interactions. A field experiment corroborates this and shows how indirect interactions can generate an inverse relationship between the mean and variance of a consumer's impact on its prey. When a species' effects are highly variable in sign and magnitude, they may average to seem weak over broad scales but be strong in local contexts. Thus, what is frequently considered to be 'noise' in species interaction data may be a critical part of the signal.

Three recent empirical studies were among the first to quantify the distribution of per capita interaction strengths among a suite of consumer-prey interactions using a common metric⁶⁻⁸ (see Methods). As in many empirical studies, species effects were identified as 'weak' when the removal or addition of a species failed to cause a statistically discernible mean change in the abundance of a target species. In each of these studies, a subset of the interactions that were considered 'weak' or zero on average were extremely variable among replicates (Fig. 1). For all three studies, the largest estimated range of interaction strengths among replicates occurred for species with mean effects that were considered 'weak' or 'insignificant'. In addition, the variation among replicates for these weak interactions was of the same order of magnitude as (or greater than) the strongest mean effects observed. These empirical patterns suggest two things: (1) before dismissing 'weak' interactors as unimportant, it may be important to distinguish between those with effects that are consistently weak versus those with strong, but variable, effects that average to be weak; and (2) by identifying 'strong' interactors solely by their mean impact, we may inadvertently be emphasizing those characterized by low variance. These possibilities question the basis of conservation strategies that attempt to prioritize research and management according to mean interaction strengths of species²⁴. Therefore, it is critical to understand what mechanisms might generate high variability in the effects of weak interactors.

Weak effects of one species on another may be expected to be more variable than strong ones under some circumstances. The net effect of a species on a target species includes a suite of direct and indirect interaction pathways that often have effects of opposite sign. Thus a species may be empirically identified as weak if its direct and indirect effects on a target species counterbalance each other. If no single direct interaction dominates, indirect effects may create large variances in the net effect on the target species because of fluctuations in the abundance of other species mediating the strengths of indirect effects. By contrast, a disproportionately

strong direct effect on the target species may swamp natural variation in indirect effects and result in a consistently strong net effect.

Using a simple rocky-intertidal food web as a model system, I experimentally tested the relationship among the strength of a consumer's mean direct effect on a prey, the importance of indirect effects, and the consistency of the net impact. In contrast to the studies reviewed above, I measured 'effect strength' as a consumer's collective impact (that is, per capita effect $\times$ density) on the abundance of a target species in order to quantify more directly the consequences of a species loss in a particular situation. Rather than compare effects of different consumers, my experiments held species identity constant and manipulated effect strength as a heuristic tool to compare variability of weak and strong effects. This approach avoided confounding factors of different species having different natural histories. Although it is simple to manipulate the density component of effect strength, altering the per capita effect is less straightforward. One proposed approach for weakening the per capita effect of a predator on a target species is to add alternative prey, assuming that the predator must then trade off some of its preference for consuming the target species¹⁷. In this study, I manipulated both consumer density and the presence of an alternative prey species. However, adding the alternative prey species also added potentially important and confounding indirect effects on the target species. Therefore, I focus on the results of altering predator density as a tool for manipulating its collective effect strength on the target species and for examining the influence of indirect effects on the outcome of predation.

The study was conducted on the rocky, central coast of Oregon, USA, in mid-intertidal disturbance patches that consisted primarily of predatory whelks and two groups of sessile prey, mussels and acorn barnacles²⁵. The mid-intertidal zone was dominated by a bed of the large California mussel, *Mytilus californianus*, which is characterized by a mosaic of disturbance patches. In the first year of succession, the patch food web consisted primarily of predatory snails (whelks: *Nucella emarginata* and *N. canaliculata*) and two species of its early colonizing sessile prey (mussels: *Mytilus trossulus*; and acorn barnacles: *Balanus glandula*). Although whelks have a negative direct effect on mussels, potential indirect effects of whelks on mussels mediated through barnacles can be both positive and negative depending on the recruitment intensity of barnacles. Barnacles directly facilitate the recruitment of mussels²⁵, so whelks can have a negative indirect effect on mussels by eating barnacles. However, at high barnacle densities, some thinning of barnacles by whelk predation can strengthen the remaining barnacles' attachment to the rock, thereby having a positive indirect effect on mussels by increasing the stability of their settlement substrate. Therefore,

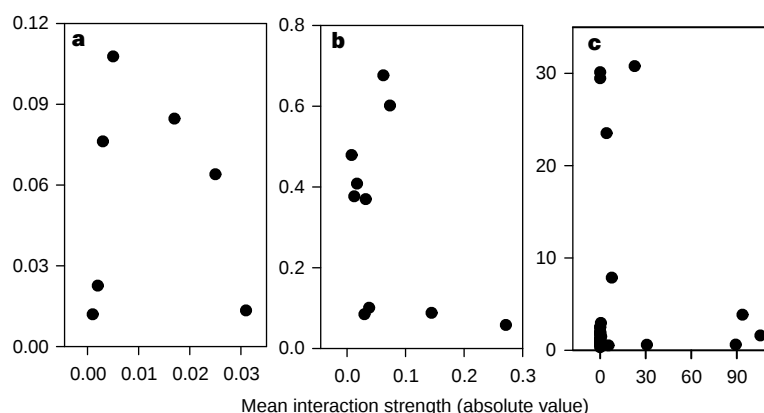


Figure 1 Estimated range of variation in interaction strength versus the mean strength for three studies that quantified *per capita* interaction strengths among a suite of consumer-prey interactions. **a**, Effects of six rocky intertidal grazers on kelp⁶; **b**, effects of preying mantids on arthropod prey⁷ (three cases were omitted

where the denominator of the interaction strength index was ≤ 0 when estimating its variation); and **c**, effects of six estuarine predators on six benthic invertebrate prey in the Ythan river estuary, Scotland⁸.

this 'simple' interaction web is ideal for testing whether and/or when indirect interactions are important in generating high variation in the effects of a consumer on its prey. When the direct effect of whelk predation on mussels is weak, its net impact should be more variable than if the direct effect is strong, because a weak direct effect would be more sensitive to variation in the strength of indirect effects mediated through barnacles.

To test this hypothesis, I conducted two experiments. The first quantified how the strength of the direct, pairwise effect of whelks on mussels varied with whelk density. In this 'mussel transplant' experiment mussels were transplanted to otherwise

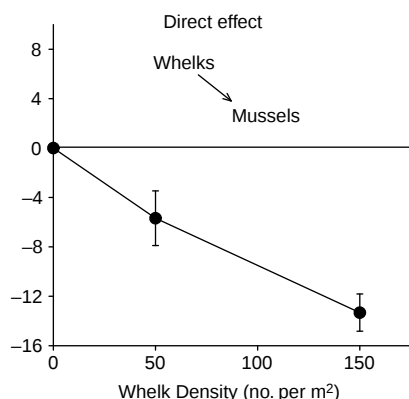


Figure 2 Mean ($\pm$ s.e.) direct, pairwise effect of whelks on mussels that were transplanted to achieve constant initial cover, for each of the three densities of whelks enclosed. One-way blocked ANOVA on the change in mussel cover over time: whelk effect significant, $P = 0.01$, d.f. = 2, $MS = 120.4$, $F = 13.0$. Linear regression of the change in mussel cover over time for the three whelk densities: $R^2 = 0.71$, $P = 0.001$.

bare plots so that all had similar initial mussel cover, and whelk density was manipulated using enclosures (see Methods). Whelks had a significant negative direct effect on mussel cover. The strength of this direct effect increased linearly with whelk density (Fig. 2) such that the per capita effects of whelks remained constant across the density of whelks manipulated. All else being equal, the density treatments successfully altered the strength of the direct, 'collective', or 'population-level' effect of whelks on mussels. Based on the results of this transplant experiment, I considered the direct, population-level effect of low- versus high-density predation on mussels to be analogous to 'weak' versus 'strong' predation, respectively, when other factors (such as the presence/absence of alternate prey) were equal.

A second 'factorial experiment' manipulated all combinations of predator density (none, low and high) and the presence of alternative prey (barnacles) (see Methods). With this design, I quantified the degree to which the net impacts of weak and strong predation on mussels were differentially sensitive to indirect effects mediated by barnacles. All treatments were replicated in each of four separate disturbance patches in the mussel bed, and the entire experiment was repeated (in separate plots in the same four patches) over three successive starting dates (1991, 1992 and 1993) which varied naturally in the recruitment of barnacles (Fig. 3 inset). This design allows the consistency of whelk impacts on mussels to be assessed in the face of spatial or temporal variation in barnacle abundance.

When barnacles were removed, both low- and high-density predation had similarly consistent negative direct effects on mussels (Fig. 3a–c). These results differ from those of the transplant experiment (Fig. 2) because mussel colonization is naturally low when barnacles are not present to facilitate recruitment²⁵. Thus, with no alternative prey, both predator treatments essentially eliminated the few mussels that settled²⁵. Note also that the whelks' direct effect (Fig. 3a–c) is consistently smaller in magnitude

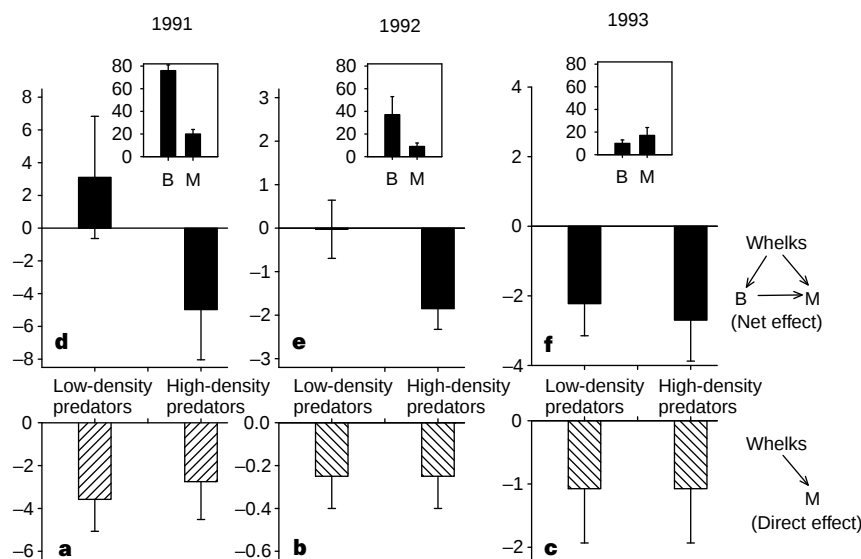


Figure 3 Mean ($\pm$ s.e., $n = 4$) effect of low- and high-density predator treatments on mussel colonization (relative to predator exclusions) for each of the three experimental starting dates. **a–c**, Newly settled barnacles were removed monthly, and thus the effects of predation include primarily direct effects on mussels (M) alone. **d–f**, Barnacles (B) were allowed to colonize, and thus the effects of predation included both direct effects and indirect effects mediated by barnacles. The effect of predator density varied with the presence of barnacles and this interaction varied among starting dates: predator density $\times$ barnacles $\times$ start date was significant: $P = 0.0001$, $F = 6.44$, d.f. = 6. The effects of low- and high-density predators did not differ significantly from each other in any of the barnacle-removal treatments (F -protected least-squares means: $P > 0.6$ in all

cases). Where barnacles were present, low- and high-density predator treatments differed significantly in 1991 and 1992, but not in 1993 (F -protected least-squares means: $P = 0.008$, 0.037 and 0.287 , respectively). Insets: Variation among years in the cumulative recruitment and mortality of barnacles and mussels in the absence of predation and interspecific competition. Data are mean ($\pm$ s.e.) covers of barnacles ($B = B. glandula$) and mussels ($M = M. Trossulus$) averaged over time for each year. Barnacle colonization varied significantly among years: ANOVA on barnacle cover in insets, effect of year: $P = 0.010$, $F = 12.79$, d.f. = 2. Mussel colonization did not vary significantly between years: ANOVA on mussel cover in insets, effect of year: $P = 0.42$, $F = 1.01$, d.f. = 2.

than the net effect of that whelk treatment (Fig. 3d–f). This is due to the fact that, even if whelks completely eliminated mussels in the absence of barnacles, because fewer mussels settle when barnacles are absent (compared to when they are present), the difference in mussel cover between treatments with and without predators is necessarily small.

When barnacles were present, low- and high-density treatments had qualitatively different net effects on mussel colonization. Strong (that is, high-density) predation was insensitive to the presence of barnacles and continued to have negative net effects on mussel colonization for all three starting dates (Fig. 3d–f). However, the mean net effect of weak predation varied in sign between starting dates and was directly related to annual variation in barnacle abundance (Fig. 3d–f). It ranged from weakly positive in 1991 to zero in 1992 to negative in 1993. Averaged over all three starting dates, the net effect of weak predation was close to zero (a change of 0.28% mussel cover per month) and thus statistically not significant. However, the range of variation among years in the mean net effect of weak predation was greater in magnitude than the largest mean net effect observed under strong predation (Fig. 3d–f: mean net effect of weak predation ranges from +3.1 in 1991 to –2.2 in 1993 versus mean net effect of strong predation of –5.0 in 1991).

In treatments where indirect effects due to barnacles were present, weak predation magnified variation in mussel colonization rates relative to no predation, whereas strong predation had a dampening effect (Fig. 4). This pattern was consistent both among replicates within a given experimental starting date as well as among starting dates (Fig. 4). The only case when weak predation reduced variation in mussel colonization relative to no predation was in 1993 when barnacle recruitment was low and their associated indirect effects unimportant (Figs 4 and 3c, f and inset).

Therefore, in this simple intertidal interaction web, when the direct effect of predation was weak, the net effect varied more in sign and magnitude than when predation was strong. All else being equal, the mean net effect of weak predation on mussels was more sensitive than strong predation to indirect effects mediated by barnacles. This resulted in greater variability in mussel colonization, both spatially among replicates and temporally among experimental starting dates, for weak predation than either no or strong predation. More important, the standard deviation of the weak predation effects equalled or exceeded the mean effect size of strong predation. Thus, although the net effect of ‘weak’ predation was not statistically significant on average, it was ecologically important in

generating variation among replicate patches and among years in mussel colonization rates.

Because the mid-intertidal mussel bed is characterized by a mosaic of disturbance patches, a species’ impact that varies greatly among patches may cause divergence of early successional trajectories, and thus increase among-patch heterogeneity in community structure²⁵. For example, in the present study individual replicate plots with barnacles under weak predation diverged rapidly in species composition. Eight months after the 1991 experimental disturbance, these weak predation replicates ranged from domination by barnacles (72% *B. glandula*; 17% *M. Trossulus*) to domination by mussels (28% *B. glandula*; 68% *M. Trossulus*). This pattern is consistent with other observations that this system is typified by disturbance patches of the same age differing dramatically in species composition^{25,26}.

More generally, these and other results^{15,17,21} challenge the assumption that research and management should focus solely on species that exhibit strong mean impacts on community structure (such as keystone species^{14,24}). In doing so, they broaden our understanding of the consequences of species loss for community organization. First, the total effect of deleting a species includes both density and per capita effects, and the effect of a species when rare can be qualitatively different from that of the same species when abundant^{27,28}. Second, although species effects that are consistently weak may have important stabilizing effects on communities¹⁷, a subset of species effects that are weak on average may be characterized by high variance. The latter can be due to a greater influence of indirect interactions when the direct effect is weak. Conditions that can generate high variation in the effect of a consumer on a target prey include when direct and indirect effects are opposite in sign, and when the balance of direct versus indirect effects shifts easily with variation in the abundance of alternative prey. Because indirect interactions are widespread in natural communities, ‘weak’ (but strongly variable) interactors may play an important, but unappreciated, role in maintaining landscape-scale diversity if their effects on species abundances are strongly context-dependent, or highly variable over space and time. A critical challenge for ecologists is thus to evaluate more generally the conditions under which weak interactions magnify rather than dampen variation in natural communities, and to understand the consequences of variable species impacts for community and ecosystem organization. □

Methods

Analysis of published interaction strength data. These three studies^{6–8} were chosen for analysis because they used a common metric and were the only ones of this type where estimates of variation in the interaction strength index can be obtained from the published data. All three studies measured per capita interaction strength using an index proposed by Paine⁶, which is expressed as: $(E - C)/(C \times P)$, where E and C are prey abundance in treatments where predators are present and absent, respectively, and P is predator density. Although per capita effects are not the same as the total impact of deleting a species, I chose to use the metric that was originally used when they distinguished between strong and weak interactors. Paine⁶ provided bootstrapped estimates of standard error (s.e.) and the number of replicates for each grazer–algae interaction strength measured. In this case, a potential range of variation for each interaction was estimated as the experimentally determined mean interaction strength ± 1 s.d. Fagan and Hurd⁷ provided s.e. for prey abundance in each treatment (C and E), but not for the interaction strength index. In this case, a range of variation was estimated by recalculating the index using $(C + 1 \text{ s.d. and } E - 1 \text{ s.d.})$ and $(C - 1 \text{ s.d. and } E + 1 \text{ s.d.})$. Raffaelli and Hall⁸ provided 95% confidence intervals for prey abundance in the C and E treatments. In this case, a range of variation was estimated as for Fagan and Hurd’s study, but with confidence intervals instead of s.d.

Mussel transplant experiment. Small mussels (<2.5 cm shell length) were transplanted to 20 × 20 cm plots that were scraped of macroscopic organisms (see ref. 23 for transplanting technique). Mean initial cover of mussels in all plots was $57 \pm 2\%$, and the change in mussel cover was monitored after 6

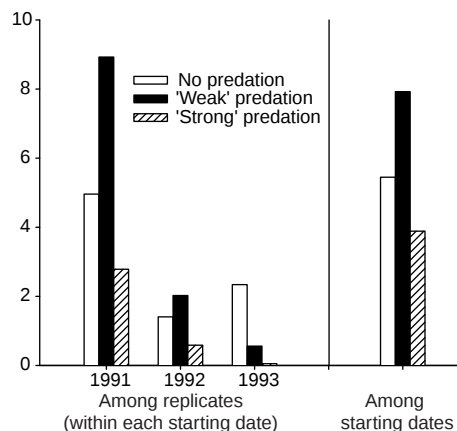


Figure 4 Variation in mussel colonization rates (per cent cover per month) between individual replicates ($n = 4$ disturbance patches) within each separate experimental starting date and between means of each starting date. Data are standard deviations of colonization rates for each predator density in treatments where barnacles were allowed to colonize.

months, when mussel cover in some replicates approached zero. In each plot, whelk densities were manipulated using 20 cm × 20 cm stainless steel cages attached to the rock. Cages either excluded whelks or enclosed two or six whelks corresponding to a range of densities naturally observed (0, 50 and 150 whelks per m², respectively). All treatments were replicated in four large patches in the mid-intertidal mussel bed. One high-density replicate was lost because of winter storm damage. Whelk effects were quantified as the difference in mussel colonization rates (change in per cent cover per month) between –whelk treatments and +whelk treatments (either low or high density). Although the results and conclusions were qualitatively identical when effects were measured using Pain's⁶ index (data not shown), I present effect strengths in terms of the difference in colonization rates because this measure, unlike Pain's, does not assume equilibrium prey abundance^{16,29}.

Factorial experiment. All plots were initially scraped bare and both mussels and barnacles colonized naturally. Whelks were enclosed at three different densities (0, 50 and 150 whelks per m²) in 20 cm × 20 cm cages, and for each whelk density, barnacles were removed monthly from half the cages. All treatments were initiated in April over three successive years: 1991, 1992 and 1993. The interactive effects of predator density (none, low, high), barnacles (present, absent) and start date (1991, 1992, 1993) on mussel colonization rate was analysed using a randomized block ANOVA. Repeated measures was not used because, although all three experimental runs were initiated over 3 years in the same blocks, individual plot localities differed between years. Data (per cent cover per month) were arcsine (square root)-transformed for analysis.

Received 18 November 1998; accepted 8 February 1999.

- Paine, R. T. A note on trophic complexity and community stability. *Am. Nat.* **103**, 91–93 (1969).
- Paine, R. T. Intertidal community structure. Experimental studies on the relationship between a dominant competitor and its principal predator. *Oecologia* **15**, 93–120 (1974).
- Power, M. E. *et al.* Challenges in the quest for keystones. *BioScience* **46**, 609–620 (1996).
- Menge, B. A., Berlow, E. L., Blanchette, C., Navarrete, S. A. & Yamada, S. B. The keystone species concept: variation in interaction strength in a rocky intertidal habitat. *Ecol. Monogr.* **64**, 249–286 (1994).
- Estes, J. A. & Palmisano, J. F. Sea otters: their role in structuring nearshore communities. *Science* **185**, 1058–1060 (1974).
- Paine, R. T. Food-web analysis through field measurement of per capita interaction strength. *Nature* **355**, 73–75 (1992).
- Fagan, W. F. & Hurd, L. E. Hatch density variation of a generalist arthropod predator: population consequences and community impact. *Ecology* **75**, 2022–2032 (1994).
- Raffaelli, D. G. & Hall, S. J. in *Food Webs: Integration of Patterns and Dynamics* (eds Polis, G. & Winemiller, K.) 185–191 (Chapman and Hall, New York, 1995).
- Wootton, J. T. Estimates and tests of per-capita interaction strength: diet, abundance, and impact of intertidally-foraging birds. *Ecol. Monogr.* **67**, 45–64 (1997).
- Hooper, D. U. & Vitousek, P. M. The effects of plant composition and diversity on ecosystem processes. *Science* **277**, 1302–1305 (1997).
- Tilman, D. *et al.* The influence of functional diversity and composition on ecosystem processes. *Science* **277**, 1300–1308 (1997).
- Wardle, D. A., Zackrisson, O., Hörnberg, G. & Gallet, C. The influence of island area on ecosystem properties. *Science* **277**, 1296–1299 (1997).
- Symstad, A. J., Tilman, D., Wilson, J. & Knops, J. M. H. Species loss and ecosystem functioning: effects of species identity and community composition. *Oikos* **81**, 389–397 (1998).
- Bengtsson, J., Jones, J. & Setälä, H. The value of biodiversity. *Trends Ecol. Evol.* **12**, 334–336 (1997).
- McGrady-Steed, J., Harris, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).
- Berlow, E. L., Navarrete, S. A., Briggs, C. J., Power, M. E. & Menge, B. A. Quantifying variation in the strengths of species interactions. *Ecology* (in press).
- McCann, K., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).
- Polis, G. A. & Strong, D. R. Food web complexity and community dynamics. *Am. Nat.* **147**, 813–846 (1996).
- Chapin, F. S., Lubchenco, J. & Reynolds, H. J. in *Global Biodiversity Assessment, United Nations Environmental Programme* (eds Heywood, V. H. & Watson, R. T.) Chapter 5.2.2 (Cambridge Univ. Press, Cambridge, 1995).
- Lawton, J. H. & Brown, V. K. in *Biodiversity and Ecosystem Function* (eds Schultze, E. D. & Mooney, H. A.) 255–270 (Springer, Berlin, 1993).
- Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).
- Frost, T. M., Carpenter, S. R., Ives, A. R. & Kratz, T. K. *Linking Species and Ecosystems* (eds Jones, C. G. & Lawton, J. H.) 224–239 (Chapman and Hall, New York, 1995).
- Navarrete, S. A. & Menge, B. A. Keystone predation and interaction strength: interactive effects of predators on their main prey. *Ecol. Monogr.* **66**, 409–429 (1996).
- Westman, W. E. Managing for biodiversity: unresolved science and policy questions. *Bioscience* **40**, 26–33 (1990).
- Berlow, E. L. From canalization to contingency: historic effects in a successional rocky intertidal community. *Ecol. Monogr.* **67**, 435–460 (1997).
- Paine, R. T. & Levin, S. A. Intertidal landscapes: disturbance and the dynamics of pattern. *Ecol. Monogr.* **51**, 145–178 (1981).
- Power, M. E. Resource enhancement by indirect effects of grazers: Armored catfish, algae, and sediment. *Ecology* **71**, 897–904 (1990).
- Navarrete, S. A. Variable predation: effects of whelks on a mid intertidal successional community. *Ecol. Monogr.* **66**, 301–321 (1996).
- Laska, M. S. & Wootton, J. T. Theoretical concepts and empirical approaches to measuring interaction strength. *Ecology* **79**, 461–476 (1998).

Acknowledgements. I thank K. Buzzard, B. Daley, J. and T. Panek, B. Profit and S. Suskie for assistance in the field, B. and A. King for the use of their front yard, and G. Allison, C. Blanchette, F. S. Chapin,

J. Corbin, C. M. D'Antonio, J. Dunne, S. Hobbie, D. Hooper, B. Langford, H. Lefcort, J. Levine, J. Lubchenco, N. D. Martinez, B. McCune, B. Menge, S. Navarrete, R. Paine, I. Parker, M. Power, W. Riley, A. Smyth, P. van Tamelen and J. T. Wootton for invaluable advice during the study. This work was supported in part by the Oregon State University Zoology Research Fund, the Holt Marine Education Fund, Sigma Xi, NSF and the Andrew W. Mellon Foundation. It was also greatly facilitated by the Sierra Nevada Aquatic Research Laboratory and the 1997–98 El Niño.

Correspondence and requests for materials should be addressed to E.L.B. (e-mail: berlow@socrates.berkeley.edu).

Anticipation of moving stimuli by the retina

Michael J. Berry II, Iman H. Brivanlou, Thomas A. Jordan* & Markus Meister

Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

A flash of light evokes neural activity in the brain with a delay of 30–100 milliseconds¹, much of which is due to the slow process of visual transduction in photoreceptors^{2,3}. A moving object can cover a considerable distance in this time, and should therefore be seen noticeably behind its actual location. As this conflicts with everyday experience, it has been suggested that the visual cortex uses the delayed visual data from the eye to extrapolate the trajectory of a moving object, so that it is perceived at its actual location^{4–7}. Here we report that such anticipation of moving stimuli begins in the retina. A moving bar elicits a moving wave of spiking activity in the population of retinal ganglion cells. Rather than lagging behind the visual image, the population activity travels near the leading edge of the moving bar. This response is observed over a wide range of speeds and apparently compensates for the visual response latency. We show how this anticipation follows from known mechanisms of retinal processing.

Because a moving object often follows a smooth trajectory, one can extrapolate from its past position and velocity to obtain an estimate of its current location. Recent experiments on motion perception^{5–7} indicate that the human brain possesses just such a mechanism: Subjects were shown a moving bar sweeping at constant velocity; a second bar was flashed briefly in alignment with the moving bar. When asked what they perceived at the time of the flash, observers reliably reported seeing the flashed bar trailing behind the moving bar. This flash lag effect has been confirmed repeatedly^{8–10}, and various high-level processes have been invoked to explain it, such as a time delay due to the shift of visual attention. To assess whether processing in the retina contributes to this effect we analysed the 'neural image' of these two stimuli at the retinal output. We recorded simultaneously the spike trains of many ganglion cells in the isolated retina of tiger salamander or rabbit. The responses to flashed and moving bars were then analysed by plotting the firing rate in the retinal ganglion-cell population as a function of space and time.

Figure 1 illustrates the responses of individual OFF-type ganglion cells to a dark bar flashed briefly over the receptive-field centre. In both salamander (Fig. 1a) and rabbit (Fig. 1b), the cells remained silent for a latency of ~50 ms, then fired a burst of spikes that lasted another 50 ms. When the bar was swept over the retina at constant speed (Fig. 1c, d), these same cells fired for a more extended period, beginning some time before the bar reached the position at which the flash occurred, and extending for a shorter time thereafter. When the bar was swept in the opposite direction (Fig. 1e, f), it produced a very similar response, showing that these cells had no direction-selective preference.

* Present address: Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California 94305, USA.

From many single-unit measurements such as these, we compiled the neural image of the visual stimulus in the population of ganglion cells. This was done by plotting the firing rate of every cell as a function of distance from the bar stimulus and interpolating these points with a smooth line (see Methods). The neural image of the flashed bar among salamander 'fast OFF' cells is shown in Fig. 2a. After a latency of 40 ms, a hump of neural activity appears that increases rapidly to a peak at 60 ms, then declines and disappears at 100 ms. As might be expected, the profile is centred on the bar. It has a width on the retina of $\sim 200 \mu\text{m}$ at half-maximum, close to the size of the receptive-field centre for these neurons¹¹. The width increases somewhat during the late phase of the response, creating the impression of an outward 'splash'¹².

The neural image of the moving bar is shown in Fig. 2b. Again, a hump of firing activity is observed, which now sweeps over the retina along with the moving bar. If this response were subject to the same time delay as the flash, one would expect the neural image to trail behind the visual image of the bar. Instead, the hump of firing activity is clearly ahead of the centre of the bar, and the peak firing rate seems to occur near the bar's leading edge. The same response occurs when the bar moves in the opposite direction. By superposing on this the response to a flash that was aligned with the moving bar (from Fig. 2a), we find that the two neural images are clearly separated—at the time when the response to the flash peaks, the response of the moving bar is displaced $\sim 100 \mu\text{m}$ ahead in the direction of motion. A very similar displacement between the two neural images was observed among brisk-sustained OFF cells in the rabbit retina (Fig. 2c). If subsequent stages of the visual system estimate the location of the flashed bar and the moving bar by the position of these humps of neural activity, they must conclude that the moving bar is ahead of the flashed bar.

How does this apparent anticipation of the moving bar come about? One suspects that cells ahead of the bar start firing early (Figs 1c–f, 2b, c) when the bar begins to invade their receptive-field

centre. The firing profile does not extend to an equal distance behind the trailing edge of the bar, perhaps because these ganglion cells have transient responses: They fire while stimulation increases as the bar invades the receptive field, not while stimulation decreases as the bar leaves it. However, we found that spatial and temporal filtering by the ganglion cell's receptive field was by itself insufficient to explain the response profiles (see below). Instead, there is another important component, which was revealed in experiments varying the intensity of the moving bar.

Figure 3a illustrates the response of this neural population to dark bars of increasing contrast relative to the background. As expected, bars of higher contrast produced stronger modulations in firing. The peak firing rate increased in proportion to contrast at first, but then appeared to saturate (Fig. 3a, inset). In addition, the shape of the neural image changed significantly with contrast. At low contrast, the peak in firing occurred behind the bar's leading edge. At high contrast—the same condition as in Fig. 2—the peak of the profile was ahead of the leading edge, followed by a more gradual decline in firing. The saturation of the peak firing rate and the shift in the response profile can be explained if the high-contrast stimulus somehow desensitizes the response of the ganglion cell after a short

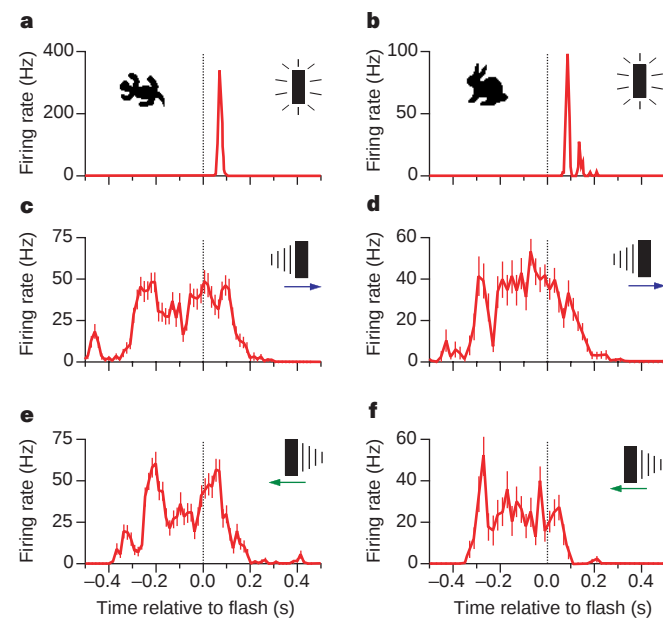


Figure 1 Responses of two ganglion cells to flashed and moving bars. **a, c, e** Results from a 'fast OFF' ganglion cell in salamander retina; **b, d, f**, results from a brisk-sustained OFF cell in rabbit retina. **a, b**, Firing rate as a function of time after a dark bar (90% contrast, $133 \mu\text{m}$ width) was flashed for 15 ms on the receptive-field centre. **c, d**, Firing rate of the same two cells while the dark bar moved continuously across the retina at 0.44 mm s^{-1} . At time zero, the bar was aligned with the position of the flash in **a** and **b**. **e, f**, As in **c, d**, but with the bar moving in the opposite direction. Error bars denote standard error across repeated presentations of the stimulus.

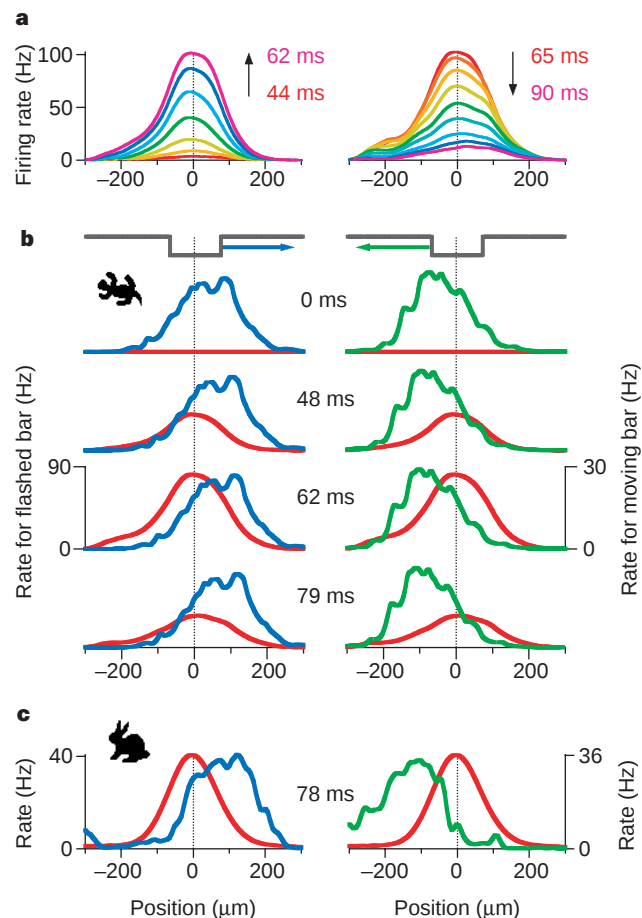


Figure 2 Population response to flashed and moving bars. **a**, Spatial profile of firing in the population of salamander fast OFF ganglion cells in response to a flashed dark bar (90% contrast, $133 \mu\text{m}$ width, see stimulus trace in **b**) at a series of times after the flash (colour scale, 3 ms steps). **b**, Profile of the population response at four time points following a flashed bar (red, from **a**), and the same bar travelling at 0.44 mm s^{-1} rightward (blue) and leftward (green). At time 0 ms, the moving bars were aligned with the position of the flash; at 62 ms, the flash response was maximal. Curves in **a** and **b** derived from 15 cells. **c**, As in **b**, for the population of brisk-sustained OFF cells in rabbit retina. At 78 ms, the flash response was maximal. Curves derived from six cells.

time delay. In that case, a ganglion cell just ahead of the bar should be strongly excited as the edge begins to enter its receptive-field centre, but then its response gain gets reduced and the firing rate declines even before the edge is half-way across. A well-known component of retinal processing that fits this description is the 'contrast-gain control'^{13–15}.

Following ref. 16, we incorporated this aspect into a quantitative description of a ganglion cell's light response (Fig. 4). In this scenario, the retina integrates the light stimulus over space and time, with a weighting function $k(x, t)$ given by the ganglion cell's receptive field, and the resulting signal determines the neuron's firing rate. If the stimulus provides strong excitation for an extended period of time, a negative feedback loop reduces the gain at the input and consequently the response to subsequent stimulation¹⁷. With just four free parameters, this model produced a satisfying account of neural responses throughout the entire contrast series (Fig. 3a). It indicates that the retinal gain is modulated as much as fourfold during passage of the high-contrast bar (Fig. 3b), which pushes the response profile towards the leading edge of the bar. Without contrast-gain control the predicted profile always lagged significantly behind (Fig. 3a).

This explanation indicates that there will be clear limits to what stimuli can be anticipated. For example, if the bar moves fast enough to cross the receptive field before the contrast-gain control sets in, then the peak of the firing profile should lag behind the leading edge. Figure 5a explores these limits, and shows that up to speeds of about 1 mm s^{-1} on the retina, the shape of the firing profile among ganglion cells remained essentially unchanged, with a peak near or ahead of the leading edge. At higher speeds, however, the response profile began to slip significantly behind the leading edge.

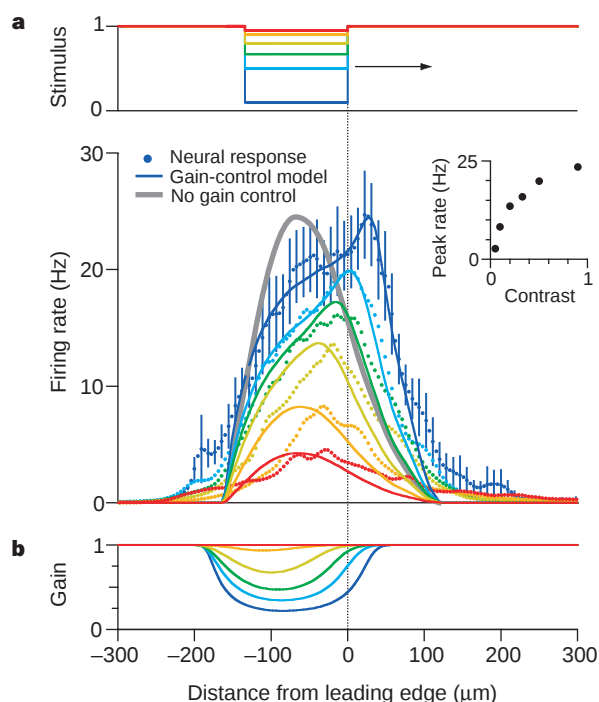


Figure 3 Dependence of motion extrapolation on contrast. **a**, Stimulation with moving dark bars ($133 \mu\text{m}$ width, 0.44 mm s^{-1} speed) of varying contrast: 5, 10, 20, 33, 50 and 90% (see top stimulus traces). Main panel shows the response profile derived from 15 salamander fast OFF ganglion cells (coloured dots), and the predicted response (coloured lines) from a model incorporating contrast-gain control (Fig. 4). Grey line shows the prediction without contrast-gain control. Inset, the peak firing rate of the response profile as a function of the contrast of the bar. Error bars denote standard error, derived from variation among ganglion cells. **b**, Gain variable g of the gain control model (Fig. 4). Model parameters: $\theta = 0$, $\alpha = 85 \text{ Hz}$, $B = 45 \text{ s}^{-1}$, $\tau = 170 \text{ ms}$.

This basic relationship was confirmed for several different populations of ganglion cells in both rabbit and salamander (Fig. 5b). The various cell types differed in the extent of anticipation at low speeds, but all began to show a lag in the neural image at speeds of $1\text{--}2 \text{ mm s}^{-1}$. In particular, direction selectivity does not play a special role in motion anticipation.

In summary, we have shown that the extrapolation of a moving object's trajectory begins in the retina. In the neural image that the eye transmits to the brain, the moving object is clearly ahead of the corresponding flashed object (Fig. 2). According to a successful model for the ganglion cell's light response (Figs 3, 4), motion anticipation in these populations can be explained on the basis of the spatially extended receptive field, the biphasic temporal response and a nonlinear contrast-gain control. There are several indications that this retinal mechanism contributes strongly to human perception of moving stimuli. First, the requisite components of our model are well documented in many species. In the primate retina, a nonlinear contrast-gain control is found specifically in the M-type ganglion cells¹⁵, neurons that feed the central pathways leading to motion perception¹⁸. Second, retinal motion extrapolation breaks down at speeds above 1 mm s^{-1} (Fig. 5b). This corresponds well with observations on human subjects: At retinal speeds of $0.3\text{--}0.9 \text{ mm s}^{-1}$, perceptual motion extrapolation appeared to compensate for the entire visual delay⁵, whereas at speeds of $\sim 4 \text{ mm s}^{-1}$ only partial extrapolation was observed¹⁰. Finally, the retina anticipates high-contrast stimuli more than low-contrast stimuli (Fig. 3), a further departure from ideal extrapolation. Again, this effect has been observed in human psychophysics⁹.

In general, an animal is likely to benefit from anticipating the

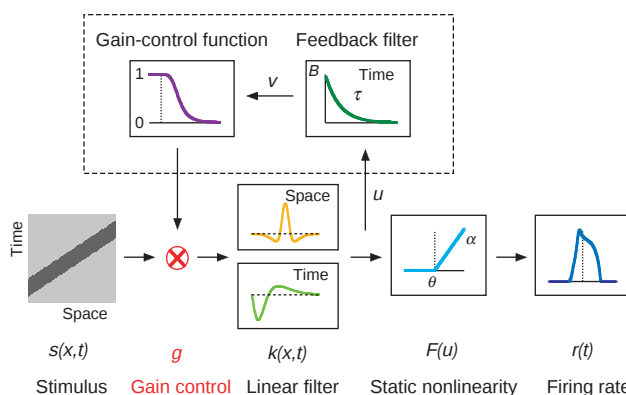


Figure 4 Cascade model for a ganglion cell's light response. The stimulus $s(x, t)$ is multiplied by a gain factor g , convolved with a spatiotemporal filter $k(x, t)$, and rectified by a static nonlinear function $F(u)$ to produce the firing rate $r(t)$. The contrast-gain control mechanism (boxed region) takes the output of the linear filter u , averages it by exponential filtering in time, and uses the result v to set the gain factor g through a decreasing gain control function $g(v)$. Formally,

$$u(t) = g(v) \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dt' s(x, t') k(x, t - t')$$

$$v(t) = \int_{-\infty}^{\infty} dt' u(t') B \exp\left(-\frac{t - t'}{\tau}\right)$$

$$g(v) = \begin{cases} 1 & v < 0 \\ 1/(1 + v^4) & v > 0 \end{cases}$$

$$F(u) = \begin{cases} 0 & u < \theta \\ \alpha(u - \theta) & u > \theta \end{cases}$$

The filter $k(x, t)$ is measured (see Methods), and $g(v)$ is taken from a previous, successful model of contrast-gain control in the salamander retina¹⁷. Thus, the model has four parameters: the threshold θ and slope α of the rectifier $F(u)$, and the amplitude B and time constant τ of the gain control filter.

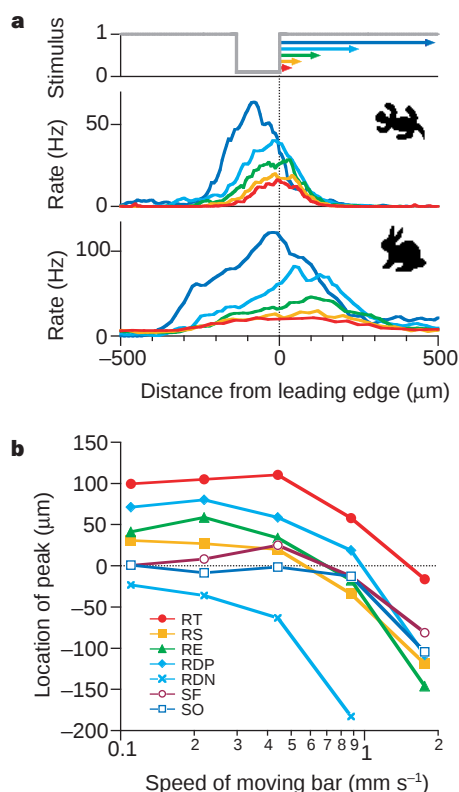


Figure 5 Dependence of motion extrapolation on speed. **a**, Stimulation with moving dark bars (90% contrast, $133 \mu\text{m}$ width) of varying speed: 0.11, 0.22, 0.44, 0.88 and 1.76 mm s^{-1} (see top stimulus traces). Firing profiles are plotted for salamander fast OFF ganglion cells (middle panel, 19 cells) and rabbit brisk-transient OFF cells (bottom panel, 3 cells). **b**, Motion extrapolation as a function of speed for various populations of ganglion cells. The distance between the peak in the firing-rate profile and the leading edge of the moving bar is plotted as a function of speed; positive numbers indicate that the response profile peaked ahead of the leading edge. The functional types are: salamander, fast OFF (SF, 19 cells); salamander, other OFF (SO, 16 cells); rabbit, brisk-transient OFF (RT, 3 cells); rabbit, brisk-sustained OFF (RS, 6 cells); rabbit, local edge detectors (RE, 4 cells); rabbit, ON/OFF direction-selective cells (5 cells) probed in the preferred direction (RDP) and the null direction (RDN).

future position of an object, for example to pounce on it or to evade it. This is particularly urgent when the primary sensory data are delayed. In principle, this delay could be compensated anywhere within the behavioural loop, even within the motor system that executes the response. However, it is advantageous to perform the correction early, before different sensory pathways merge. For example, in many animals the retina projects directly to the tectum or the superior colliculus, where a visual map of space is overlaid with an auditory map^{19,20}. Auditory transduction in hair cells incurs a much shorter delay than phototransduction²¹. If the visual and auditory images of a moving object should align on the target map, the compensation for the delay in the visual pathway must occur within the retina.

It is likely that subsequent stages of the visual system continue this process, possibly by using a similar mechanism. Within the visual cortex we can certainly find the requisite components of the model in Fig. 4: for example, local pooling of excitatory inputs, time-delayed inhibition and mechanisms of nonlinear gain control^{22,23}. More generally, there are many instances within the cortex where variables relevant to our behaviour are mapped onto two-dimensional sheets of neurons²⁴. If the time course of these variables produces a smooth trajectory of neural activity on the cortical map, then a mechanism such as that described here can predict their future from past observations. □

Methods

Recording. Retinae were obtained from larval tiger salamanders and Dutch belted rabbits. A piece of isolated retina was placed ganglion-cell-layer-down on a multi-electrode array, which recorded spike trains simultaneously from many ganglion cells, as described previously^{11,25}.

Stimulation. Visual stimuli were generated on a computer monitor and projected onto the photoreceptor layer, as described²⁵. All experiments used a background of white light, with a photopic intensity of $M = 11 \text{ mW m}^{-2}$. Dark bars of intensity B were presented on this background, and the contrast of a bar is defined as $C = (M - B)/M$. A screen pixel of the monitor measured $6.7 \mu\text{m}$ on the retina, and each video frame lasted 15 ms. Thus, a bar sweeping at 0.44 mm s^{-1} moved by one pixel every video frame. Flashed bars were presented for a single video frame.

Receptive fields. The spatiotemporal receptive fields of all ganglion cells were measured by reverse correlation to randomly flickering stripes²⁵, orientated parallel to the bars from other experiments. Each of the contiguous $13\text{-}\mu\text{m}$ wide stripes was randomly turned on or off every 30 ms. From ~ 60 min of recording, we computed for each ganglion cell the average stimulus sequence in the one second preceding an action potential. This reverse correlation is a measure of how the ganglion cell integrates light over space and time. Its time-reverse is the ganglion cell's linear kernel $k(x, t)$ (ref. 26), which can also be interpreted as the effect of a thin line flashed at distance x on the cell's firing rate at time t after the flash¹¹. As expected, $k(x, t)$ had a 'Mexican hat' spatial profile, reflecting opposite effects from centre and surround²⁷, and a biphasic time course (see Fig. 4 and ref. 11).

Cell types. Retinal ganglion cells appear in distinct functional types, and we took care to analyse these subpopulations separately. Salamander cells were classified based on their spatiotemporal receptive fields and on responses to uniform square-wave flashes, as described²⁸. Rabbit cells were classified based on the spatiotemporal receptive field and the shape of the spike train's autocorrelation function, following the criteria of ref. 29. Direction-selective cells produced at least tenfold more spikes to one direction of the moving bar than to the opposite direction.

Population activity. The profile of population activity was evaluated along the spatial dimension perpendicular to the bars. Each cell's position was defined as the middle of its receptive field, determined by fitting a spatial gaussian to the centre lobe of the kernel $k(x, t)$. To estimate the population response to a flashed bar, the flash was repeated 75 times in each of 15 locations, separated by $33 \mu\text{m}$. For each flash location and each ganglion cell, the firing rate following the flash (Fig. 1a, b) was calculated using a time bin of 2 ms. To compose the firing profile at a given time after the flash, each ganglion cell's firing rate was plotted against the cell's position relative to the flashing line. This plot was smoothed by convolution with a gaussian of standard deviation $20 \mu\text{m}$. To estimate the population response to a moving bar, the stimulus was repeated 50 times, and each ganglion cell's firing rate computed as for flashes, but with a time bin of 15 ms. Then the firing rate was plotted against the distance from the bar, and averaged over all cells.

Received 11 December 1998; accepted 2 February 1999.

- Maunsell, J. H. & Gibson, J. R. Visual response latencies in striate cortex of the macaque monkey. *J. Neurophysiol.* **68**, 1332–1344 (1992).
- Lennie, P. The physiological basis of variations in visual latency. *Vision Res.* **21**, 815–824 (1981).
- Schnapf, J. L., Kraft, T. W. & Baylor, D. A. Spectral sensitivity of human cone photoreceptors. *Nature* **325**, 439–441 (1987).
- De Valois, R. L. & De Valois, K. K. Vernier acuity with stationary moving Gabors. *Vision Res.* **31**, 1619–1626 (1991).
- Nijhawan, R. Motion extrapolation in catching. *Nature* **370**, 256–257 (1994).
- Khurana, B. & Nijhawan, R. Extrapolation or attention shift? *Nature* **378**, 566 (1995).
- Nijhawan, R. Visual decomposition of colour through motion extrapolation. *Nature* **386**, 66–69 (1997).
- Baldo, M. V. & Klein, S. A. Extrapolation or attention shift? *Nature* **378**, 565–566 (1995).
- Purushothaman, G., Patel, S. S., Bedell, H. E. & Ogmen, H. Moving ahead through differential visual latency. *Nature* **396**, 424 (1998).
- Whitney, D. & Murakami, I. Latency difference, not spatial extrapolation. *Nature Neurosci.* **1**, 656–657 (1998).
- Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W. & Meister, M. Adaptation of retinal processing to image contrast and spatial scale. *Nature* **386**, 69–73 (1997).
- Jacobs, A. L. & Werblin, F. S. Spatiotemporal patterns at the retinal output. *J. Neurophysiol.* **80**, 447–451 (1998).
- Shapley, R. M. & Victor, J. D. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol.* **285**, 275–298 (1978).
- Sakai, H. M., Wang, J. L. & Naka, K. Contrast gain control in the lower vertebrate retinas. *J. Gen. Physiol.* **105**, 815–835 (1995).
- Benardete, E. A., Kaplan, E. & Knight, B. W. Contrast gain control in the primate retina: P cells are not X-like, some M cells are. *Visual Neurosci.* **8**, 483–486 (1992).
- Victor, J. D. The dynamics of the cat retinal X cell centre. *J. Physiol.* **386**, 219–246 (1987).

17. Crevier, D. W. & Meister, M. Synchronous period-doubling in flicker vision of salamander and man. *J. Neurophysiol.* **79**, 1869–1878 (1998).
18. Merigan, W. H. & Maunsell, J. H. How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* **16**, 369–402 (1993).
19. Sparks, D. L. Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus. *Physiol. Rev.* **66**, 118–171 (1986).
20. Knudsen, E. I. Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* **2**, 1177–1194 (1982).
21. Corey, D. P. & Hudspeth, A. J. Response latency of vertebrate hair cells. *Biophys. J.* **26**, 499–506 (1979).
22. Carandini, M., Heeger, D. J. & Movshon, J. A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17**, 8621–8644 (1997).
23. Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. Synaptic depression and cortical gain control. *Science* **275**, 220–224 (1997).
24. Knudsen, E. I., du Lac, S. & Esterly, S. D. Computational maps in the brain. *Annu. Rev. Neurosci.* **10**, 41–65 (1987).
25. Meister, M., Pine, J. & Baylor, D. A. Multi-neuronal signals from the retina: acquisition and analysis. *J. Neurosci. Methods* **51**, 95–106 (1994).
26. Hunter, I. W. & Korenberg, M. J. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biol. Cybern.* **55**, 135–144 (1986).
27. Rodieck, R. W. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Res.* **5**, 583–601 (1965).
28. Warland, D. K., Reinagel, P. & Meister, M. Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.* **78**, 2336–2350 (1997).
29. Devries, S. H. & Baylor, D. A. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J. Neurophysiol.* **78**, 2048–2060 (1997).

Acknowledgements. We thank J. Keat for assistance in generating the visual stimulus, and H. Berg and T. Holy for comments on the manuscript. This work was supported by a NRS to M.B. and a grant from the NIH and a Presidential Faculty Fellowship to M.M.

Correspondence and requests for material should be addressed to M.J.B. (e-mail: berry@biosun.harvard.edu).

A new cellular mechanism for coupling inputs arriving at different cortical layers

Matthew E. Larkum, J. Julius Zhu & Bert Sakmann

Abt. Zellphysiologie, Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, D-69120 Heidelberg, Germany

Pyramidal neurons in layer 5 of the neocortex of the brain extend their axons and dendrites into all layers. They are also unusual in having both an axonal and a dendritic zone for the initiation of action potentials^{1–6}. Distal dendritic inputs, which normally appear greatly attenuated at the axon, must cross a high threshold at the dendritic initiation zone to evoke calcium action potentials^{1,7} but can then generate bursts of axonal action potentials. Here we show that a single back-propagating sodium action potential generated in the axon⁸ facilitates the initiation of these calcium action potentials when it coincides with distal dendritic input within a time window of several milliseconds. Inhibitory dendritic input can selectively block the initiation of dendritic calcium action potentials, preventing bursts of axonal action potentials. Thus, excitatory and inhibitory postsynaptic potentials arising in the distal dendrites can exert significantly greater control over action potential initiation in the axon than would be expected from their electrotonically isolated locations. The coincidence of a single back-propagating action potential with a subthreshold distal excitatory postsynaptic potential to evoke a burst of axonal action potentials represents a new mechanism by which the main cortical output neurons can associate inputs arriving at different cortical layers.

Triple recordings were made on two sites of the apical dendrites and the somata of layer-5 pyramidal neurons (Fig. 1a). Subthreshold excitatory-postsynaptic-potential (EPSP)-shaped potentials in the distal apical dendrites were strongly attenuated as they spread to the soma (Fig. 1b). Back-propagating action potentials from the axon reached the distal dendritic tufts (Fig. 1c) where they caused an influx⁹ of Ca^{2+} . However, single sodium action potentials (Na^+ -APs) initiated in the axon do not evoke a calcium action potential (Ca^{2+} -AP) in the dendrite^{1,3,4,6–9} (Fig. 1c). The combination of a subthreshold EPSP-shaped distal dendritic potential

(Fig. 1b) and a back-propagating action potential (Fig. 1c) elicited a Ca^{2+} and Na^+ action potential complex in the dendrite (Fig. 1d; $n = 27$) that resulted in 2–3 (2.2 ± 0.1 ; frequency: 101 ± 8 Hz) Na^+ -APs at the soma. This back-propagating action potential activated Ca^{2+} spike firing (BAC firing) was evoked with 1.1 nA (± 0.1 nA) peak current injection at a distal dendritic location (average $693 \pm 18 \mu\text{m}$ from soma). This was only half the current amplitude ($49 \pm 4\%$; $n = 27$) that evoked a comparable Ca^{2+} -AP in the absence of a back-propagating action potential (Fig. 1e). In the neuron shown in Fig. 1d, the current needed was only 0.3 nA (25% of threshold). The threshold dendritic current for BAC firing was typically less than the threshold for eliciting a Na^+ -AP using the same-shaped current injection at the soma, suggesting that the dendritic component requires fewer synaptic inputs than are needed for axonal Na^+ -APs (possibly less than ten inputs¹).

The optimal time for eliciting a Ca^{2+} -AP with dendritic current injection was 3–7 ms after the somatic action potential. At threshold, the time interval (Δt) for evoking BAC firing was very narrow. Here, dendritic current injection given at Δt values of 3, 7 and 11 ms (Fig. 2a–c, respectively) elicited BAC firing only at $\Delta t = 7$ ms. The average threshold for BAC firing was measured at 5-ms time intervals (Fig. 2d; $n = 8$) and revealed a sharp minimum at $\Delta t = 5$ ms. The threshold for BAC firing rose sharply for delays of over 5 ms, and was even higher than the threshold for eliciting a Ca^{2+} -AP in the absence of a back-propagating action potential for delays of 10–130 ms. This shows that the generation of a Ca^{2+} -AP is greatly facilitated if the dendritic EPSP and axonal action potential

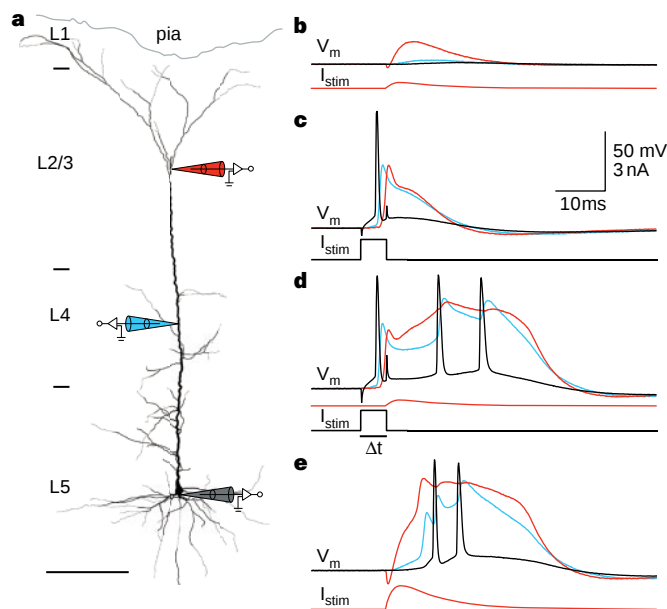


Figure 1 Coupling of a back-propagating action potential (AP) with distal subthreshold current injection. **a**, Reconstruction of a biocytin-filled pyramidal neuron, with the recording pipette positions shown symbolically (770 μm from soma in red, 400 μm from soma in blue and one at the soma in grey). Cortical layers are indicated on the left. Scale bar, 200 μm . **b**, Current injection of 0.3 nA (peak amplitude) at the distal pipette (red trace, bottom) in the shape of an EPSP produced a signal of only 1.4 mV at the soma. It did not reach threshold for either a Ca^{2+} -AP or a Na^+ -AP. I_{stim} refers to traces representing current injected and V_m to potential recorded. Positive current causes intracellular depolarization. The colour indicates the corresponding electrode in the diagram. **c**, Threshold current injection at the soma (black trace labelled I_{stim}) evoked an AP that reduced in amplitude but increased in width in the dendrite. **d**, BAC firing. The combination of these injections of current (used in **b** and **c**) separated by an interval (Δt) of 5 ms evoked a burst of APs following the onset of the Ca^{2+} -AP in the distal dendrite. (Δt , time between the onset of each pulse.) Scale bars in **c** also apply to **b** and **d**. **e**, A similar dendritic Ca^{2+} -AP could be evoked by a larger (1.2 nA) current injection alone at the distal dendritic electrode, as shown previously¹.

are timed so that they coincide within a few milliseconds, but is slightly depressed if the EPSP follows a back-propagating action potential 10–130 ms later.

The mechanism underlying the generation of bursts of action potentials is the fact that a Ca^{2+} -AP, once it is elicited in the dendritic initiation zone, can cause enough depolarization at the axonal initiation zone to cross threshold in adult (>P28) rats⁷ (Fig. 1e). It gives rise to complex dendritic potentials once threshold has been reached. The back-propagating action potential substantially lowers the threshold for the initiation of this dendritic Ca^{2+} -AP. This event then initiates a burst of axonal action potentials that propagate back into the dendritic arbor. Therefore the final output pattern of the neuron, when combining a back-propagating action potential with a subthreshold EPSP, consists of the first action potential that is generated in the axon, and the second and third action potentials that are an intrinsic property of the neuron for suprathreshold distal dendritic input. This even occurred in cells in which prolonged somatic depolarization did not elicit such a burst pattern. BAC firing was found in all 27 layer-5 pyramidal neurons tested, of which 21 were 'regularly spiking' and 6 were 'intrinsically bursting', as defined by somatic current injection¹⁰.

Extracellularly evoked subthreshold synaptic input (Fig. 3a, b),

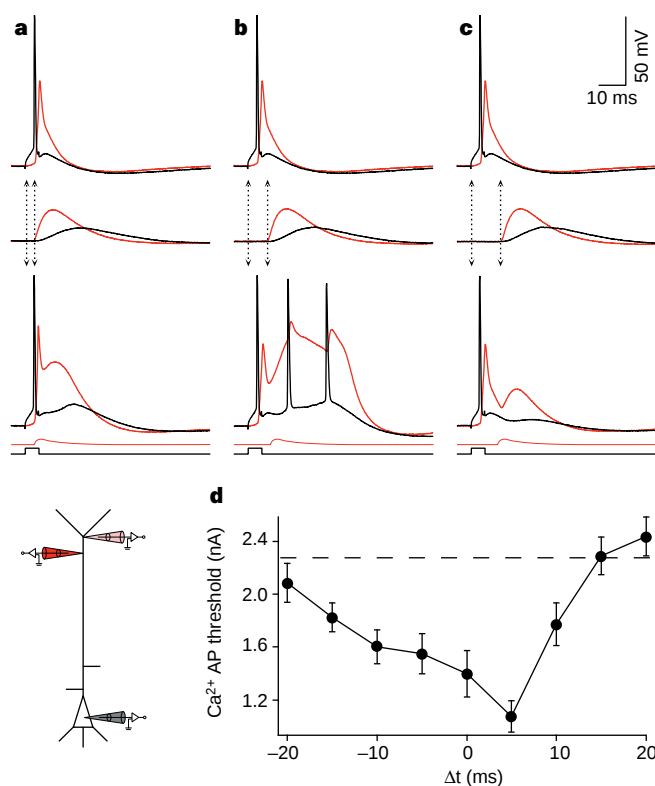


Figure 2 Precision of timing required. Experimental configuration shown diagrammatically (lower left). Recordings were made from the dendrite (red; 600 μm from the soma) and the soma (black) of a L5 pyramidal neuron. A third dendritic electrode (pink; 700 μm from the soma) was used for injecting current. (Electrode colours correspond to recording traces.) Time intervals (**a**, 3 ms; **b**, 7 ms; and **c**, 11 ms) elicited a burst of APs only in **b** at threshold. Δt was taken as the time between the start of the somatic current injection and that of the dendritic current injection. Note, however, that the AP due to the somatic current injection followed the onset by ~ 3 ms in this case. **d**, A burst of APs could be generated by the combination of dendritic current injection and a back-propagating AP at other times, but the threshold for this was least at $\Delta t = 5$ ms. Each point is the average of 8 neurons (error bars, s.e.m.) and represents the threshold for current injection needed to elicit a dendritic Ca^{2+} -AP. Dashed line represents the Ca^{2+} -AP threshold without a back-propagating AP (2.28 ± 0.14 nA). For ~ 100 ms after $\Delta t = 10$ ms, the threshold was even slightly higher than without the back-propagating AP.

coupled with a back-propagating action potential, could generate a dendritic Ca^{2+} -AP (in 3 of 10 neurons; Fig. 3c). In most neurons (7 out of 10), the Ca^{2+} -AP was not as large or was non-existent until the addition of low concentrations of GABA antagonists, bicuculline (1 μM) and saclofen (100 μM), which increased the proportion of excitatory to inhibitory input. After application, extracellularly evoked synaptic input coupled with a back-propagating action potential could always evoke a Ca^{2+} -AP in the dendrite ($n = 3/3$). A relatively small extracellularly evoked EPSP was sufficient (an EPSP of 9 mV in the distal dendrite was the smallest needed). Besides showing that synaptic input could generate BAC firing, this experiment also implied that inhibition might selectively modulate the generation of the Ca^{2+} -AP. It is known, for instance, that GABAergic inhibition can affect dendritic Ca^{2+} -APs and Ca^{2+} influx due to back-propagating action potentials in hippocampal pyramidal neurons^{11–13}.

To test whether GABAergic inhibition affected the Ca^{2+} -AP component of BAC firing, we made simultaneous recordings from interneurons and dendritic and somatic recordings from target pyramidal neurons (Fig. 4a). First, BAC firing was generated in the usual way by coupling a back-propagating action potential with a subthreshold EPSP-shaped dendritic current injection (Fig. 4b). The effect of unitary inhibitory postsynaptic potentials (IPSPs) on the back-propagating action potential itself was small and was only visible when the IPSP and the back-propagating action potential were coincident. In contrast, unitary IPSPs abolished the Ca^{2+} -AP and the associated action potential burst completely (Fig. 4c, recording 1; $n = 5/5$). Trains of presynaptic action potentials in interneurons or long depolarization causing bursts of action potentials could prevent the Ca^{2+} -AP component up to 150 ms in advance (data not shown). By using an extracellular stimulus electrode positioned within 500 μm from the pyramidal neuron and by blocking excitatory transmission with 20 μM CNQX and 50 μM AP5, BAC firing was blocked up to 400 ms in advance ($n = 14$; data not shown).

The action of inhibition was located at the dendritic initiation zone. This was concluded from experiments in which the Ca^{2+} -AP was isolated from the soma. In six experiments, the Ca^{2+} -AP component of BAC firing did not generate a burst of Na^+ -APs at the soma. The Ca^{2+} -AP could also be isolated with hyperpolarizing

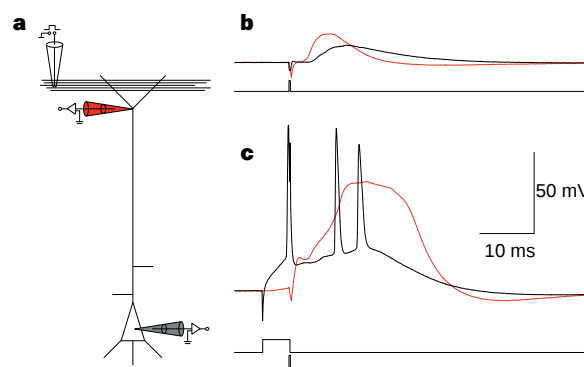


Figure 3 Extracellularly evoked BAC-firing. **a**, Experimental configuration. A bipolar electrode was placed over the layer-1 fibres ~ 500 μm from the pyramidal cell. **b**, Subthreshold extracellular stimulus of 5 V evoked an EPSP of 17.5 mV in the tufts of a L5 pyramidal neuron (800 μm from the soma; red) and 10 mV at the soma (black). Here the extracellular stimulus presumably evoked synaptic inputs arising mostly in the distal dendrites but also partly near the soma, judging from the small difference in amplitudes at both electrodes. **c**, By preceding the extracellular stimulus with a back-propagating AP ($\Delta t = 5$ ms), a Ca^{2+} -AP was evoked that was similar to the Ca^{2+} -AP evoked with direct dendritic current injection and a back-propagating AP in the same neuron. (Timing and shape of current injections shown below voltage recordings; red for the dendritic electrode and black for the somatic electrode.)

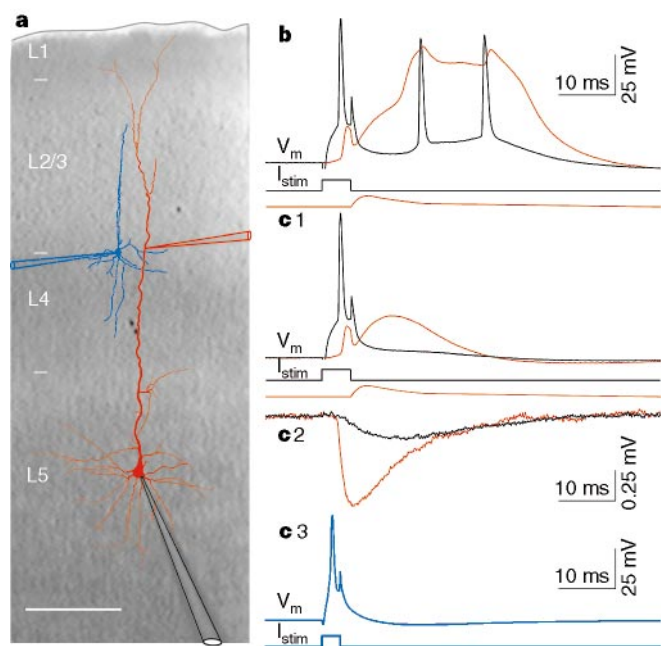


Figure 4 Inhibition of BAC firing. **a**, Reconstruction of the dendritic arbor of a biocytin-filled pyramidal neuron (red neuron) and an inhibitory neuron connected to it (blue neuron) superimposed on a phase-contrast image taken during the experiment showing the cortical layers (labels in white). Pipette positions on the dendrite (red, 550 μm from soma), soma (black) and on the presynaptic neuron (blue) are superimposed symbolically. Scale bar, 200 μm . **b**, BAC firing. Threshold was 0.7 nA when injected through the dendritic pipette 5 ms after the start of injection of current at the soma. **c**, Recordings: 1, unitary inhibitory input blocked the burst of APs but not the somatic AP; 2, the IPSP evoked by a single presynaptic AP was larger at the dendritic electrode (-0.75 mV) than at the soma (-0.2 mV); 3, AP in the presynaptic inhibitory neuron that was responsible for the unitary IPSP shown in recording 2.

current injected at the soma ($n = 2$) or by using shorter dendritic current injections (time to peak 0.1–3.2 ms; $n = 16$). In such conditions, the effect of inhibition, which abolished the isolated Ca^{2+} -AP but not the back-propagating Na^{+} -AP, must have been on the dendritic initiation zone. Moreover, Ca^{2+} -APs evoked by supra-threshold dendritic depolarization could be blocked up to 400 ms in advance ($n = 28$; data not shown) by extracellularly evoked inhibition in the presence of 50 μM AP5 and 20 μM CNQX. Hyperpolarizing or depolarizing the dendrite and/or soma immediately preceding the generation of a Ca^{2+} -AP had no effect on the threshold for evoking Ca^{2+} -APs, indicating that the action of the inhibition was not mediated through a voltage-dependent mechanism. We suggest that the mechanism might be due to shunt by GABA_A and GABA_B conductances and may also involve the activation or inactivation of dendritic conductances¹⁴.

The time-sensitive coupling of a back-propagating action potential with distal excitatory input in L5 pyramidal neurons suggests a new mechanism by which neocortical pyramidal neurons can associate synaptic inputs arriving in the upper and lower layers within a few milliseconds. Which particular input pathways might be associated will depend on the local cortical architecture and is likely to vary in different areas of the cortex. At the cellular level, the mechanism itself appears to be very robust (observed in all 27 L5 pyramidal neurons tested). It involves the co-activation of two initiation zones, which only occurs with specific spatial and temporal input.

Thus, BAC firing provides a potential mechanism for binding different cortical regions¹⁵. For example, it has been suggested that temporal binding occurs as a result of coincident activation of

specific thalamic input to layer 4 and nonspecific thalamic input to layers 1 and 2 (ref. 16). Absence of such nonspecific input is accompanied by cortical dysfunction¹⁷. Furthermore, extracellular recordings made from the somatosensory cortex of monkeys have shown that natural touch stimulation evokes a fast signal in the lower cortical layers, followed by a slow signal in the upper cortical layers¹⁸. The slow component was abolished by slow-wave-sleep-induced inhibition.

BAC firing alone would presumably be unstable in an interconnected structure like the cortex and lead to cells continuously bursting in the absence of some other regulatory mechanism. We have shown that inhibition can selectively modulate the ability of the neuron to fire a Ca^{2+} -AP without blocking the generation of Na^{+} -APs due to depolarization at the axonal initiation site. Inhibition, therefore, would act to decouple inputs that would otherwise be associated. Surprisingly, the effect of inhibition appears to last for up to ~ 400 ms, in contrast to the relatively precise timing of the back-propagating action potential and distal EPSP required for BAC firing to occur. This suggests that inhibition acts as a form of general veto on this kind of burst firing.

Ca^{2+} -APs occur *in vivo* owing to synaptic input^{5,12,19,20}. As these Ca^{2+} -APs can cause bursts of action potentials that might signify the association of functionally separated input, it is crucial to understand what sort of postsynaptic effects a burst in L5 pyramidal neurons may have and what sort of information it might encode²¹. Of course, bursts may also occur for reasons other than BAC firing. Our results show that BAC firing could serve as a mechanism for facilitating such spiking behaviour.

In summary, at least two regions of L5 pyramidal neurons can generate action potentials. One has a low threshold and integrates predominantly input at basal and apical oblique dendrites, and the other integrates mostly input that arrives in the distal regions of the apical dendrite. If threshold is reached at the axonal site of integration, the subsequent back-propagating action potential enables threshold at the dendritic initiation site to be reached more easily. Thus, near-synchronous input to both distal and proximal dendrites causes a burst of axonal action potentials, whereas the same-sized input to either region alone does not cause a burst. The complex interaction between the dendritic and axonal initiation sites described here therefore constitutes a new cellular mechanism by which the firing behaviour of output neurons of the cortex can be modulated by synchronous synaptic inputs arriving at different lamina. Furthermore, inhibition can selectively modulate the coupling of the two initiation zones without reducing the ability of the cell to discharge axonal action potentials. □

Methods

Experiments were done in somatosensory neocortical slices from P28–P58 Wistar rats ($n = 57$). Rats were deeply anaesthetized by halothane and decapitated. The brain was quickly removed into cold (0 – 4°C) oxygenated physiological solution containing (in mM): NaCl, 125; KCl, 2.5; NaH_2PO_4 , 1.25; NaHCO_3 , 25; MgCl_2 , 1; dextrose, 25; and CaCl_2 , 2, pH 7.4. Sagittal slices, 250–300 μm thick, were cut from the tissue blocks with a microslicer and kept at 37°C .

All experiments were done at $34.0 \pm 0.5^\circ\text{C}$, with dual and triple recordings from single, identified layer-5 pyramidal neurons using infrared illumination combined with differential interference contrast optics²². Somatic (5–10 M Ω) and dendritic (10–25 M Ω) recording pipettes were filled with the standard intracellular solution containing (in mM): potassium gluconate, 115; HEPES, 10; MgCl_2 , 2; Mg-ATP , 2; Na_2ATP , 2; GTP, 0.3; KCl, 20; and biocytin, 0.25%, pH 7.3. For some recordings, 30 mM KCl and 105 mM potassium gluconate were used. Whole-cell recordings were made with up to three Axoclamp-2B amplifiers (Axon Instruments). Dendritic current injection was made in the shape of a double exponential ($f(t) = (1 - e^{-t/\tau_1}) \times e^{-t/\tau_2}$) with τ_1 being 0.5–2 ms and τ_2 , 2–8 ms. The function was adjusted so that the peak current injection corresponded to the value referred to as the amplitude.

Extracellularly evoked synaptic stimulation was induced by a bipolar

electrode with single-voltage pulses (100–200 μ s, up to 5 V). The stimulating electrode was typically placed over layer 1 about 500 μ m lateral to the cells. After recordings, slices were fixed and stained as described¹. All results are reported as mean $\pm$ s.e.m.

Received 23 October 1998; accepted 25 January 1999.

- Schiller, J., Schiller, Y., Stuart G. & Sakmann, B. Calcium action potentials restricted to the distal apical dendrites of rat neocortical pyramidal neurons. *J. Physiol.* **505**, 605–616 (1997).
- Spencer, W. A. & Kandel, E. R. Electrophysiology of hippocampal neurons IV. Fast prepotentials. *J. Neurophys.* **24**, 272–295 (1961).
- Magee, J., Hoffman, D., Colbert, C. & Johnston, D. Electrical and calcium signaling in dendrites of hippocampal pyramidal neurons. *Annu. Rev. Physiol.* **60**, 327–346 (1998).
- Kim, H. G. & Connors, B. W. Apical dendrites of the neocortex: correlation between sodium- and calcium-dependent spiking and pyramidal cell morphology. *J. Neurosci.* **13**, 5301–5311 (1993).
- Hirsch, J. A., Alonso, J. M. & Reid, C. R. Visually evoked calcium action potentials in cat striate cortex. *Nature* **378**, 612–616 (1995).
- Rhodes, P. A. Functional Implications of Active Currents in the Dendrites of Pyramidal Neurons in *Cerebral Cortex* (eds Uliniski, P. & Jones, E. G.) Vol. 13, 139–200 (Plenum, New York, 1999).
- Zhu, J. J. & Sakmann, B. Postnatal development of Ca^{2+} -mediated action potentials in dendritic tufts of rat neocortical pyramidal neurons. *Soc. Neurosci. Abstr.* **23**, 2283 (1997).
- Stuart, G. & Sakmann, B. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature* **367**, 69–72 (1994).
- Markram, H., Helm, P. J. & Sakmann, B. Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. *J. Physiol. (Lond.)* **485**, 1–20 (1995).
- Connors, B. W. & Gutnick, M. J. Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.* **13**, 99–104 (1990).
- Miles, R., Toth, K., Gulyas, A. I., Hajos, N. & Freund, T. F. Differences between somatic and dendritic inhibition in the hippocampus. *Neuron* **16**, 815–823 (1996).
- Buzsáki, G., Penttonen, M., Nádasdy, Z. & Bragin, A. Pattern and inhibition-dependent invasion of pyramidal cell dendrites by fast spikes in the hippocampus *in vivo*. *Proc. Natl Acad. Sci. USA* **93**, 9921–9925 (1996).
- Tsubokawa, H. & Ross, W. N. IPSPs modulate spike backpropagation and associated $[Ca^{2+}]_i$ changes in the dendrites of hippocampal CA1 pyramidal neurons. *J. Neurophys.* **76**, 2896–906 (1996).
- Chen, H. & Lambert, N. A. Inhibition of dendritic calcium influx by activation of G-protein-coupled receptors in the hippocampus. *J. Neurophysiol.* **78**, 3484–3488 (1997).
- Singer, W. & Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
- Llinás, R., Ribary, U., Joliot, M. & Wang, X.-J. in *Temporal Coding in the Brain* (eds Buzsáki, G. et al.) 251–272 (Springer, Berlin, 1994).
- Heilman, K. M. & Valenstein, E. *Clinical Neuropsychology* (Oxford Univ. Press, New York, 1993).
- Cauller, L. J. & Kulics, A. T. The neural basis of the behaviorally relevant N1 component of the somatosensory-evoked potential in SI cortex of awake monkeys: evidence that backward cortical projections signal conscious touch sensation. *Exp. Brain Res.* **84**, 607–618 (1991).
- Zhu, J. J. & Sakmann, B. Whisker-evoked slow oscillation (7–12 Hz) in single neurons of the rat barrel cortex. *Soc. Neurosci. Abstr.* **24**, 1512 (1998).
- Zhu, J. J. & Connors, B. W. Intrinsic firing patterns and whisker-evoked synaptic responses of neurons in the rat barrel cortex. *J. Neurophysiol.* **81**, 1171–1183 (1999).
- Lisman, J. E. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci.* **20**, 38–43 (1997).
- Dodt, H. U. Infrared-interference videomicroscopy of living brain slices. *Adv. Exp. Med. Biol.* **333**, 245–249 (1993).

Acknowledgements. We thank B. Katz, A. Korngreen, G. Borst and A. Silver for their helpful comments. M.E.L. was supported by an Alexander von Humboldt scholarship and J.J.Z. was supported by a Max-Planck-Gesellschaft fellowship.

Correspondence and requests for materials should be addressed to M.E.L. (e-mail: mlarkum@sunny.mpmf-heidelberg.mpg.de).

Signalling through CD30 protects against autoimmune diabetes mediated by CD8 T cells

Christian Kurts*, Francis R. Carbone†, Matthew F. Krummel‡, Karl M. Koch*, Jacques F. A. P. Miller‡ & William R. Heath‡

* The Department of Nephrology, Medizinische Hochschule, 30625 Hannover, Germany

† The Department of Pathology and Immunology, Monash Medical School, Prahran, Victoria 3181, Australia

‡ The Immunology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia

Autoantigens found on pancreatic islets can move to draining lymph nodes, where they are able to cause the activation and consequent deletion of autoreactive T cells by a mechanism termed cross-tolerance^{1,2}. This deletion depends on signalling through CD95 (also known as Fas), a member of the superfamily

of tumour-necrosis-factor receptors³. Here we describe a new mechanism that protects against autoimmunity: this mechanism involves another member of this superfamily, CD30, whose function was largely unknown. CD30-deficient islet-specific CD8-positive T cells are roughly 6,000-fold more autoaggressive than wild-type cells, with the transfer of as few as 160 CD30-deficient T cells leading to the complete destruction of pancreatic islets and the rapid onset of diabetes. We show that, in the absence of CD30 signalling, cells activated but not yet deleted by the CD95-dependent cross-tolerance mechanism gain the ability to proliferate extensively upon secondary encounter with antigen on parenchymal tissues, such as the pancreatic islets. Thus, CD30 signalling limits the proliferative potential of autoreactive CD8 effector T cells and protects the body against autoimmunity.

CD30 is a member of the tumour-necrosis-factor receptor (TNFR) superfamily, and is expressed by activated, but not by resting, B or T cells^{4–8}. Although its function is largely unknown, *in vitro* studies have shown that it has effects on both cell death and cell activation^{7–11}. T cells of the T-helper-2 phenotype may express CD30 (ref. 6), although this has been questioned¹². Mice deficient in CD30 showed a mild impairment in thymic negative selection¹³, but a function for CD30 in the *in vivo* response of mature T cells has yet to be demonstrated.

Here we investigated the role of CD30 signalling in the maintenance of self-tolerance by using a model system in which ovalbumin (OVA)-specific CD8⁺ T cells from the OT-I transgenic line (OT-I cells) were adoptively transferred into unirradiated transgenic mice that expressed OVA in the pancreatic β -cells and the proximal tubular cells of the kidney (RIP-mOVA mice)¹⁴. In this model, wild-type OT-I cells caused diabetes only when adoptively transferred in relatively large numbers (>250,000 cells) (Table 1), with lower doses being effectively tolerized. In contrast, CD30-deficient OT-I cells (OT-I.CD30^{–/–} cells) were highly diabetogenic, with as few as 160 cells causing disease. The diabetogenic potential of such small doses of CD30-deficient cells indicates that CD30 is important in the prevention of autoimmunity mediated by CD8⁺ T cells.

We have shown previously that OT-I cells are slowly deleted by a CD95-dependent mechanism in the weeks following transfer into RIP-mOVA mice³. CD30 deficiency did not impair this slow deletion process (data not shown), but seemed to exert its effect much earlier than did CD95, as indicated by the rapid onset of diabetes as early as 4 days after adoptive transfer. To gain understanding of what was happening in these few days following adoptive transfer, we used flow cytometry to count wild-type and CD30-deficient OT-I cells in the spleen and lymph nodes of RIP-mOVA mice shortly after adoptive transfer (Fig. 1). There was a marked antigen-specific increase (expansion) in the number of CD30-deficient cells, but not of wild-type OT-I cells, peaking on days 4–5 after adoptive transfer into RIP-mOVA mice. This expansion was accompanied by prominent infiltration of the CD30-deficient cells into the pancreas and by β -cell destruction (data not shown). Thus, CD30-deficient cells

Table 1 CD30-deficient OT-I cells cause diabetes in RIP-mOVA mice

No. of injected cells	No. of diabetic mice/no. in group	
	OT-I	OT-I.CD30 ^{–/–}
5,000,000	18/18	5/5
2,000,000	9/10	9/9
1,000,000	4/8	6/6
250,000	0/26	9/10
100,000	0/7	9/11
20,000		12/13
4,000		6/6
800		9/13
160		4/9

Data from several experiments are pooled. Mice were injected with semi-purified OT-I or OT-I.CD30^{–/–} T cells and then examined for diabetes by analysis of urine glucose as described¹⁵. In some experiments, diabetes was confirmed by analysis of blood glucose, and islet destruction was confirmed by histology as described¹⁶. Diabetes onset occurred between days 4 and 10, depending on the number of cells transferred.

proliferated extensively in the few days following adoptive transfer, indicating that CD30 signalling may normally limit the proliferative potential of autoreactive cells.

When transferred into RIP-mOVA mice, OT-I cells respond to antigen sequentially, in two different phases, each of which occurs in a separate site. In the priming phase of the response, naïve OT-I cells are activated by OVA cross-presented on bone-marrow-derived antigen-presenting cells (APCs) in the draining lymph nodes of the kidney and pancreas; then, during the effector phase, activated OT-I cells enter parenchymal tissues and attack cells synthesizing OVA (for example, pancreatic-islet β -cells)^{1,14,15}. Under normal circumstances, OT-I cells primed in this manner generate a very weak effector phase which is able to cause diabetes only when large numbers of OT-I cells ($>250,000$) are adoptively transferred (ref. 15 and Table 1). Lower numbers of OT-I cells are not pathogenic and are eventually deleted through the CD95-dependent mechanism³, thus maintaining self-tolerance.

To determine where CD30 exerted its effect on the proliferation, we studied the response of wild-type and CD30-deficient OT-I cells in both sites of antigen recognition, that is, the pancreatic islets and the draining lymph nodes. To examine proliferation during the effector phase in the pancreatic islets, we adoptively transferred 2×10^5 wild-type or CD30-deficient OT-I cells into RIP-mOVA mice and then 5 days later injected a 6-h pulse of the thymidine analogue bromodeoxyuridine (BrdU), which is incorporated into the DNA of proliferating cells *in situ*. This short pulse was used to ensure that those BrdU-labelled cells found in the pancreas had proliferated in this site. Immunohistological examination of the pancreas revealed many BrdU-positive (proliferating) cells associated with the islets of mice given CD30-deficient OT-I cells, but no BrdU-positive cells in mice given wild-type OT-I cells (Fig. 2). This showed that CD30-deficient cells, in contrast to wild-type cells, proliferated extensively in the pancreatic islets.

Analysis of the proliferation rates of OT-I cells during the priming phase in the pancreatic draining lymph nodes revealed no difference between wild-type or CD30-deficient cells on days 2 or 3 after adoptive

transfer (data not shown). Thus, CD30 did not affect proliferation during the priming phase in the draining nodes. Taken together, these data indicate that the preferential expansion of CD30-deficient cells occurred upon recognition of antigen in the pancreatic islets.

Although we had shown that CD30-deficient cells proliferated preferentially after recognizing antigen on islet cells, it was important to determine whether priming in the draining nodes was a prerequisite for this recognition to occur; that is, was there a link between priming in the nodes and proliferation in the islets? To address this issue, we asked whether expansion of CD30-deficient OT-I cells required antigen recognition in the draining nodes (during the priming phase), in the parenchymal tissues (during the effector phase), or in both tissues. We generated chimaeric mice using two congenic mouse strains, C57Bl/6 (B6) and bm1, which differ at the major-histocompatibility-complex (MHC) locus $H-2K$. OVA can be presented to OT-I cells by the MHC class-I molecule $H-2K^b$ (expressed by B6 mice), but not by $H-2K^{bm1}$ (expressed by bm1 mice)¹⁶. After introducing bm1 bone marrow into RIP-mOVA mice of the B6 type (bm1 $\rightarrow$ B6), the bone-marrow-derived cells of the recipients were incapable of presenting OVA to OT-I cells, and thus priming in the draining lymph nodes was blocked, whereas presentation of OVA by parenchymal tissues remained functional. Likewise, introduction of B6 bone marrow into RIP-mOVA mice of the bm1 type (B6 $\rightarrow$ bm1) generated chimaeras in which antigen recognition on parenchymal tissues was impossible, but priming in the draining lymph nodes remained unaffected. Finally, generation of B6 $\rightarrow$ B6 chimaeras allowed both the presentation of antigen by the bone-marrow compartment during the priming phase in the draining nodes, and the presentation of antigen by islet β -cells to effector T cells that had migrated to the pancreas. After adoptive transfer of CD30-deficient cells, extensive expansion was seen only in B6 $\rightarrow$ B6 chimaeras, where presentation could take place in both the draining nodes and the pancreas (Fig. 3).

These data indicate first, that priming in the lymph nodes was necessary to allow expansion of CD30-deficient cells (B6 $\rightarrow$ B6 but

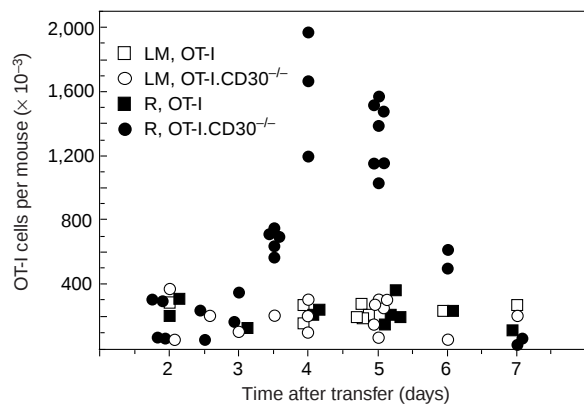


Figure 1 Expansion of CD30-deficient OT-I cells on days 4 and 5 after transfer into RIP-mOVA mice. 2×10^6 OT-I (squares) or OT-I.CD30^{-/-} (circles) cells were injected into RIP-mOVA mice (R; filled symbols) or non-transgenic littermates (LM; open symbols). At different time points, the number of OT-I cells present in the recipient mice was determined by flow cytometry after pooling their spleen and lymph-node cells and staining for expression of CD8 and the transgenic T-cell antigen-receptor chains V α 2 and V β 5 as described¹. Data from several experiments are pooled. Each symbol represents one mouse. It is usual for the recovery of cells to be significantly lower than the number adoptively transferred, as only about 10–20% of cells ‘take’ in unirradiated recipients. The decrease in the number of OT-I.CD30^{-/-} cells seen on days 6–7 in RIP-mOVA mice was associated with a general decrease in all T-cell numbers caused by the stress of acute-diabetes onset (data not shown). Of 5 RIP-mOVA mice given 2×10^6 CD30-deficient OT-I cells and followed daily for diabetes onset by testing glucosuria, all 5 became diabetic on day 5.

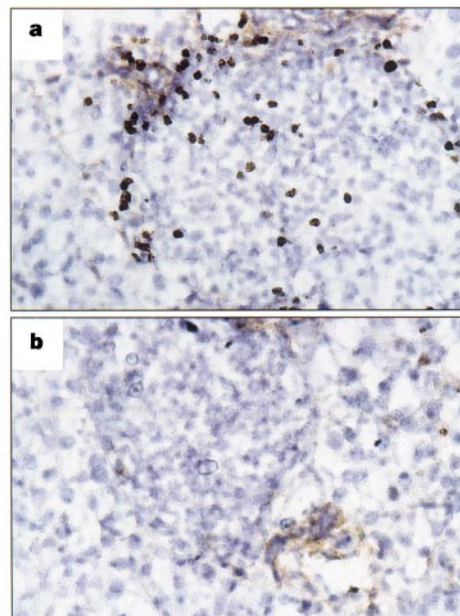


Figure 2 CD30 deficiency leads to extensive proliferation of OT-I cells in the islets of RIP-mOVA mice. RIP-mOVA mice were injected with **a**, 2×10^5 OT-I.CD30^{-/-} cells or **b**, 2×10^5 OT-I cells. Five days later, recipients were injected intraperitoneally with 1.5 mg BrdU in saline and then killed after a further 6 h. Frozen sections of pancreatic tissue were stained for BrdU incorporation as described²³. Recipients were not diabetic at the time of analysis.

not $\text{bm1} \rightarrow \text{B6}$ chimaeras showed expansion), but second, that, despite the proliferative response in the draining lymph nodes, there was no marked expansion of CD30-deficient cells unless they also recognized antigen on parenchymal cells ($\text{B6} \rightarrow \text{B6}$ but not $\text{B6} \rightarrow \text{bm1}$ chimaeras showed expansion). Thus, the protective (anti-proliferative) effect of CD30 exerts itself at the level of secondary encounter with antigen on parenchymal tissues. The requirement for priming in the draining nodes is most likely to be related to the fact that naïve T cells recirculate only within the secondary lymphoid tissues¹⁷ and will not migrate into the parenchymal tissues unless they are first activated by priming. The ability of CD30-deficient cells to proliferate in response to 'non-professional' APCs suggests that activated CD30-deficient cells have little requirement for co-stimulation by members of the B7 family.

We have shown that CD30 signalling can protect against autoimmunity by preventing extensive expansion of autoreactive CD8⁺ effector T cells during secondary encounter with antigen in parenchymal tissues. CD30 signalling may also alter other functions, such as cytokine production or expression of homing receptors, but such changes are not needed to explain our data because a simple increase in the frequency of precursor OT-I cells leads to diabetes. Analysis of the cytotoxic capacity of CD30-deficient cells *in vitro* revealed that they were no more cytotoxic than wild-type cells (data not shown). Similarly, the *in vitro* proliferative response to OVA-bearing cells showed no enhancement when cells were deficient in CD30 (data not shown). This latter result indicates that the specific characteristics of our *in vivo* model are essential to show the effect of CD30 deficiency.

Our results identify a new mechanism for control of autoimmunity that is distinct from the CD95-dependent deletion of autoreactive CD8⁺ T cells³. Thus, two levels exist for the control of such cells. The first level represents a mechanism of deletion that is dependent on CD95 and is akin to activation-induced cell death. This mechanism is rather slow, however, and results in the genera-

tion of effectors that, before their deletion, can be highly destructive. To limit the destructive potential of these effectors, the immune system uses a second level of control involving CD30, which prevents extensive proliferation of effector cells that encounter antigen on parenchymal tissues such as the islets. Interestingly CD30 signalling leads to degradation of TRAF2, a TNFR-associated factor¹⁸. Without TRAF2, signalling through TNFR-1 is biased towards the apoptotic pathway rather than the TRAF2-dependent NF- κ B-linked activation pathway. Perhaps, in the absence of CD30-mediated degradation of TRAF-2, TNF or other related molecules preferentially signal cell proliferation rather than apoptosis. It is unlikely that CD30 signalling acts indirectly through CD95 to prevent proliferation, as CD95-deficient OT-I cells are only marginally more diabetogenic than wild-type cells (data not shown), whereas very few CD30-deficient cells are required to cause diabetes (Table 1). Another alternative⁹ is that CD30 signalling may prevent proliferation by downregulating the expression of molecules such as CD28. However, as the stimulators of proliferation in our model (the islet cells) are non-professional APCs, which do not express the CD28-binding molecules B7.1 or B7.2, it is unlikely that regulation of CD28 expression is critical. However, CD30-mediated regulation of other molecules involved in cell-cell interactions could provide an explanation.

These results raise the question of why the immune system has evolved a powerful, CD30-dependent, mechanism to stop the proliferation of CD8⁺ T cells in parenchymal tissues. The cross-tolerance pathway is partner to a cross-priming pathway¹⁹ that is capable of inducing immunity to tumours, DNA vaccines and minor antigens²⁰. Perhaps CD30 signalling is suppressed under conditions of inflammation or infection, similar to the decoupling of the CD95 pathway from cross-priming that is seen in the presence of lipopolysaccharide²¹, so that activated CD8⁺ T cells can undergo an extra amplification step in the parenchymal tissue when they encounter infected cells. In contrast, for self-antigens, the CD30 signal would remain functional and severely limit the expansion of autoreactive T cells. It will be interesting to evaluate the potential of blocking CD30 signalling for the augmentation of cytotoxic-T-lymphocyte function during antitumour immunotherapy. □

Methods

Mice. C57B1/6, bm1 , RIP-mOVA¹⁴, CD30-deficient¹³ OT-I and OT-I²² mice were maintained at the Walter and Eliza Hall Institute of Medical Research.

Preparation of lymphocytes for adoptive transfer. OT-I and OT-I.CD30^{-/-} cells were prepared from the lymph nodes of transgenic mice as described¹⁴. Briefly, erythrocytes, B cells and CD4⁺ T cells were depleted by treatment with the heat-stable antigen-specific monoclonal antibody J11d and the CD4-specific monoclonal antibody RL172, followed by complement. Semi-purified cells were analysed by flow cytometry to determine the proportion of OT-I cells, and then injected intravenously into recipient mice.

Bone-marrow chimaeras. To generate bone-marrow chimaeras, 7–8-week-old mice were lethally irradiated (900 cGy) and then reconstituted with 5×10^6 T-cell-depleted bone-marrow cells. The next day, mice were further depleted of T cells by intraperitoneal injection of 100 μl of the Thy1-specific ascites, T24 ascites.

Flow-cytometric analysis. To determine the total number of OT-I cells in recipient mice, we pooled lymph-node (inguinal, brachial, axillary, mesenteric and superficial cervical) and spleen cells of recipient mice, depleted them of erythrocytes and stained them for CD8 and the OT-I T-cell-receptor chains $\text{V}\alpha 2$ and $\text{V}\beta 5$, using the monoclonal antibodies YTS169.4, B20.1 and MR9.4, respectively, as described¹. The number of OT-I cells was then calculated as (percentage $\text{CD8}^+ \text{V}\alpha 2^+ \text{V}\beta 5^+$ - 1.4) $\times$ percentage CD8^+ $\times$ total cell number, where 1.4 was the percentage of $\text{CD8}^+ \text{V}\alpha 2^+ \text{V}\beta 5^+$ T cells in a normal mouse, as described¹.

Received 17 December 1998; accepted 18 February 1999.

1. Kurts, C., Kosaka, H., Carbone, F. R., Miller, J. F. A. P. & Heath, W. R. Class I-restricted cross-presentation of exogenous self antigens leads to deletion of autoreactive CD⁸ T cells. *J. Exp. Med.* **186**, 239–245 (1997).

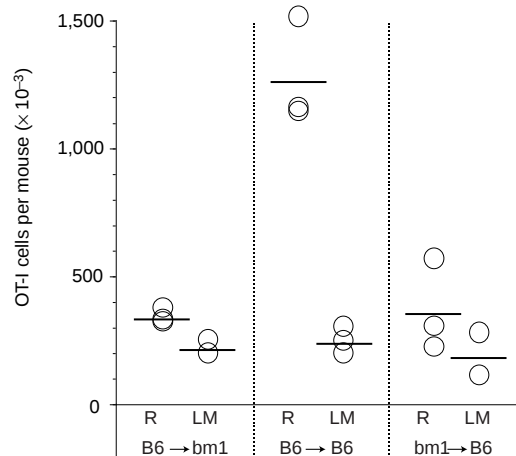


Figure 3 The expansion of autoreactive CD8⁺ T cells requires their interaction with both bone-marrow-derived APCs and parenchymal tissue cells. 2×10^6 OT-I.CD30^{-/-} were injected into chimaeric RIP-mOVA mice (R) or non-transgenic littermate control mice (LM); these chimaeric mice could present OVA only on their professional APCs ($\text{B6} \rightarrow \text{bm1}$), only on their parenchymal tissue ($\text{bm1} \rightarrow \text{B6}$), or in both compartments ($\text{B6} \rightarrow \text{B6}$). After five days, the number of OT-I cells present in the recipient mice was determined by flow cytometry as described¹. Histological analysis of pancreas sections of these chimaeric mice stained with Gomari's aldehyde fuchsin revealed massive infiltration and nearly complete β -cell destruction in the $\text{B6} \rightarrow \text{B6}$ chimaeras, but not in the $\text{B6} \rightarrow \text{bm1}$ or $\text{bm1} \rightarrow \text{B6}$ chimaeras (data not shown). Thus, expansion of CD30-deficient cells correlates with autoimmunity. Similar experiments, performed with wild-type OT-I cells, did not show cell expansion in any of the chimaeras (data not shown), confirming that expansion was specific to CD30-deficient cells.

2. Heath, W. R., Kurts, C., Miller, J. F. A. P. & Carbone, F. R. Cross-tolerance: a pathway for inducing tolerance to peripheral tissue antigens. *J. Exp. Med.* **187**, 1549–1553 (1998).
3. Kurts, C., Heath, W. R., Kosaka, H., Miller, J. F. A. P. & Carbone, F. R. The peripheral deletion of autoreactive CD8⁺ T cells induced by cross-presentation of self-antigens involves signalling through CD95 (Fas, Apo-1). *J. Exp. Med.* **188**, 415–420 (1998).
4. Schwab, U. *et al.* Production of a monoclonal antibody specific for Hodgkin and Sternberg-Reed cells of Hodgkin's disease and a subset of normal lymphoid cells. *Nature* **299**, 65–67 (1982).
5. Schwarting, R. *et al.* BER-H2: a new anti-Ki-1 (CD30) monoclonal antibody directed at a formol-resistant epitope. *Blood* **74**, 1678–1689 (1989).
6. Bowen, M. A., Lee, R. K., Miragliotta, G., Nam, S. Y. & Podack, E. R. Structure and expression of murine CD30 and its role in cytokine production. *J. Immunol.* **156**, 442–449 (1996).
7. Telford, W. G., Nam, S. Y., Podack, E. R. & Miller, R. A. CD30-regulated apoptosis in murine CD8 T cells after cessation of TCR signals. *Cell. Immunol.* **182**, 125–136 (1997).
8. Lee, S. Y., Noel, P. J., Podack, E. R., Reiner, S. L. & Thompson, C. B. Expression of the costimulatory receptor CD30 is regulated by both CD28 and cytokines. *J. Immunol.* **160**, 2180–2187 (1998).
9. Bowen, M. A., Olsen, K. J., Cheng, L., Avila, D. & Podack, E. R. Functional effects of CD30 on a large granular lymphoma cell line, YT. Inhibition of cytotoxicity, regulation of CD28 and IL-2R, and induction of homotypic aggregation. *J. Immunol.* **151**, 5896–5906 (1993).
10. Lee, S. Y. *et al.* CD30/TNF receptor-associated factor interaction: NF- κ B activation and binding specificity. *Proc. Natl Acad. Sci. USA* **93**, 9699–9703 (1996).
11. Lee, S. Y., Park, C. G. & Choi, Y. T cell receptor-dependent cell death of T cell hybridomas mediated by the CD30 cytoplasmic domain in association with tumor necrosis factor receptor-associated factors. *J. Exp. Med.* **183**, 669–674 (1996).
12. Hamann, D. *et al.* CD30 expression does not discriminate between human Th1- and Th2-type T cells. *J. Immunol.* **156**, 1387–1391 (1996).
13. Amakawa, R. *et al.* Impaired negative selection of T cells in Hodgkin's disease antigen CD30-deficient mice. *Cell* **84**, 551–562 (1996).
14. Kurts, C. *et al.* Constitutive class I-restricted exogenous presentation of self antigens *in vivo*. *J. Exp. Med.* **184**, 923–930 (1996).
15. Kurts, C. *et al.* CD4⁺ T cell help impairs CD8⁺ T cell deletion induced by cross-presentation of self-antigens and favors autoimmunity. *J. Exp. Med.* **186**, 2057–2062 (1997).
16. Nikolic, Z. J. & Carbone, F. R. The effect of mutations in the MHC class I peptide binding groove on the cytotoxic T lymphocyte recognition of the Kb-restricted ovalbumin determinant. *Eur. J. Immunol.* **20**, 2431–2437 (1990).
17. Mackay, C. R. Homing of naive, memory and effector lymphocytes. *Curr. Opin. Immunol.* **5**, 423–427 (1993).
18. Duckett, C. S. & Thompson, C. B. CD30-dependent degradation of TRAF2: implications for negative regulation of TRAF signalling and the control of cell survival. *Genes Dev.* **11**, 2810–2821 (1997).
19. Bevan, M. J. Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay. *J. Exp. Med.* **143**, 1283–1288 (1976).
20. Jondel, M., Schirmbeck, R. & Reimann, J. MHC class I-restricted CTL responses to exogenous antigens. *Immunology* **5**, 295–302 (1996).
21. Vella, A. T., McCormack, J. E., Linsley, P. S., Kappler, J. W. & Marrack, P. Lipopolysaccharide interferes with the induction of peripheral T cell death. *Immunity* **2**, 261–270 (1995).
22. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
23. Bertolino, P. *et al.* Peripheral deletion of autoreactive CD8⁺ T cells in transgenic mice expressing H-2Kb in the liver. *Eur. J. Immunol.* **25**, 1932–1942 (1995).

Acknowledgements. We thank T. Mak for the CD30-deficient mice; J. Allison for advice; and T. Banjanin, F. Karamalis and P. Nathan for technical assistance. C.K. was supported in part by a fellowship from the Deutsche Forschungsgemeinschaft. This work was funded by the NIH, the National Health and Medical Research Council of Australia and the Australian Research Council.

Correspondence and requests for materials should be addressed to W.R.H. (e-mail: Heath@wehi.edu.au) or F.R.C. (e-mail: carbone@cobra.path.monash.edu.au).

Effects of altered gene order or orientation of the locus control region on human β -globin gene expression in mice

Keiji Tanimoto, Qinghui Liu, Jörg Bungert* & James Douglas Engel

Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208-3500, USA

The five human β -type-globin genes, ϵ , $G\gamma$, $A\gamma$, δ and β , are close together and are regulated by a locus control region (LCR) located at the 5' end of the locus^{1,2}. Here we investigate the functional consequences of this organization with respect to temporal regulation of the individual genes, by using recombination techniques to invert the order of either the genes or the LCR *in vivo*. Our analysis of transgenic mice bearing either normal or mutant transgenes leads to two new observations. First, the position of the

ϵ -globin gene next to the LCR is mandatory for its expression during the yolk-sac stage of erythropoiesis. Second, LCR activity is orientation dependent, and so the LCR does not act as a simple enhancer to stimulate transcription of the globin genes. Thus, in the absence of any change in transgene integration position, transgene copy number, *trans*-acting factors or other resident genetic information, simple inversion of the human genes or the LCR fundamentally alters the transcription of β -type globin genes.

The human β -globin genes are organized in a 5'- ϵ - $G\gamma$ - $A\gamma$ - δ - β -3' array and their expression is transcriptionally controlled, resulting in different haemoglobin tetramers being produced during the embryonic (ϵ -globin), fetal (γ -globin) and adult (β -globin) stages¹. High-level expression of all of the β -like genes is regulated by the LCR, a group of five DNaseI-hypersensitive sites (Fig. 1a) found far 5' of the ϵ -globin genes of humans, galagos, rabbits, goats and mice. In these animals, an embryonic gene is always located closest to the LCR, whereas adult genes are always further away². Previous studies addressing the contribution of gene order to temporal regulation in the locus obtained the following results. First, when two equivalent genes were linked to the LCR, the proximal gene was more abundantly transcribed^{3–5}. Second, when the γ - and β -globin genes were linked to the LCR, they were regulated in appropriate stage-specific manner regardless of gene order, although their relative levels were dependent on their proximity to the LCR⁴. It thus became generally accepted that *trans*-acting factors are the primary determinants of temporal specificity of β -type-globin gene expression, whereas proximity to the LCR modulates the relative expression of the genes by altering expression ratios through promoter competition^{6,7}. However, another possibility is that gene order plays no significant part in developmental activation, and that the principal activity of the LCR is to open chromatin, thereby allowing regulatory sequences throughout the locus equal access to transcription factors⁸. Analyses of deletion mutants of individual hypersensitive sites in the LCR (either transgenes or the endogenous locus) in mice have indicated that hypersensitive sites contribute synergistically to LCR activity and, therefore, that the LCR may function as a single unit containing both strong chromatin-opening and transcriptional-stimulatory activities^{9,10}. Thus we were interested to know whether the LCR exhibits directionality.

To minimize the usual complications of analysing transgenic mice (with respect to gene copy number or effects from the transgene integration site), we used Cre-mediated *loxP* recombination^{11,12} within yeast artificial chromosome (YAC) transgenes¹³ *in vivo*. A *loxP* site was introduced, by homologous recombination in yeast, in inverted orientation between the LCR and the ϵ -globin gene promoter and also 3' of the β -globin gene enhancer, in YAC A201F4.3 (these sites were 48 kilobases (kb) apart; Fig. 1a). To modify the LCR, we placed inverted *loxP* sites 5' of hypersensitive site 5 and 3' of hypersensitive site 1 (16 kb apart; Fig. 1b). YAC DNAs were purified and injected into fertilized eggs to generate transgenic mice. Following founder identification by the polymerase chain reaction (PCR), we determined the copy number and structure of the transgenes by conventional and pulsed-field Southern blots. In the end-fragment assay¹⁴, we detected single fragments when using probes for both left and right vector arms in all the transgenic lines (lines 42 and 31 for gene inversions, and 277 and 264 for LCR inversions; Fig. 1c), indicating that all of the transgenes were both intact and single copy.

To invert the genes or LCR *in vivo*, we collected fertilized eggs produced after mating of female mice to transgenic studs, and then injected these eggs with a Cre expression plasmid. The progeny were genotyped by PCR, and Cre-mediated site-directed recombination between *loxP* sites was monitored on Southern blots (see later). Those animals (Cre F₀) in which *in vivo* recombination was detected were bred again to generate F₁ mice. To confirm the structure of the

* Present address: Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610-0245, USA.

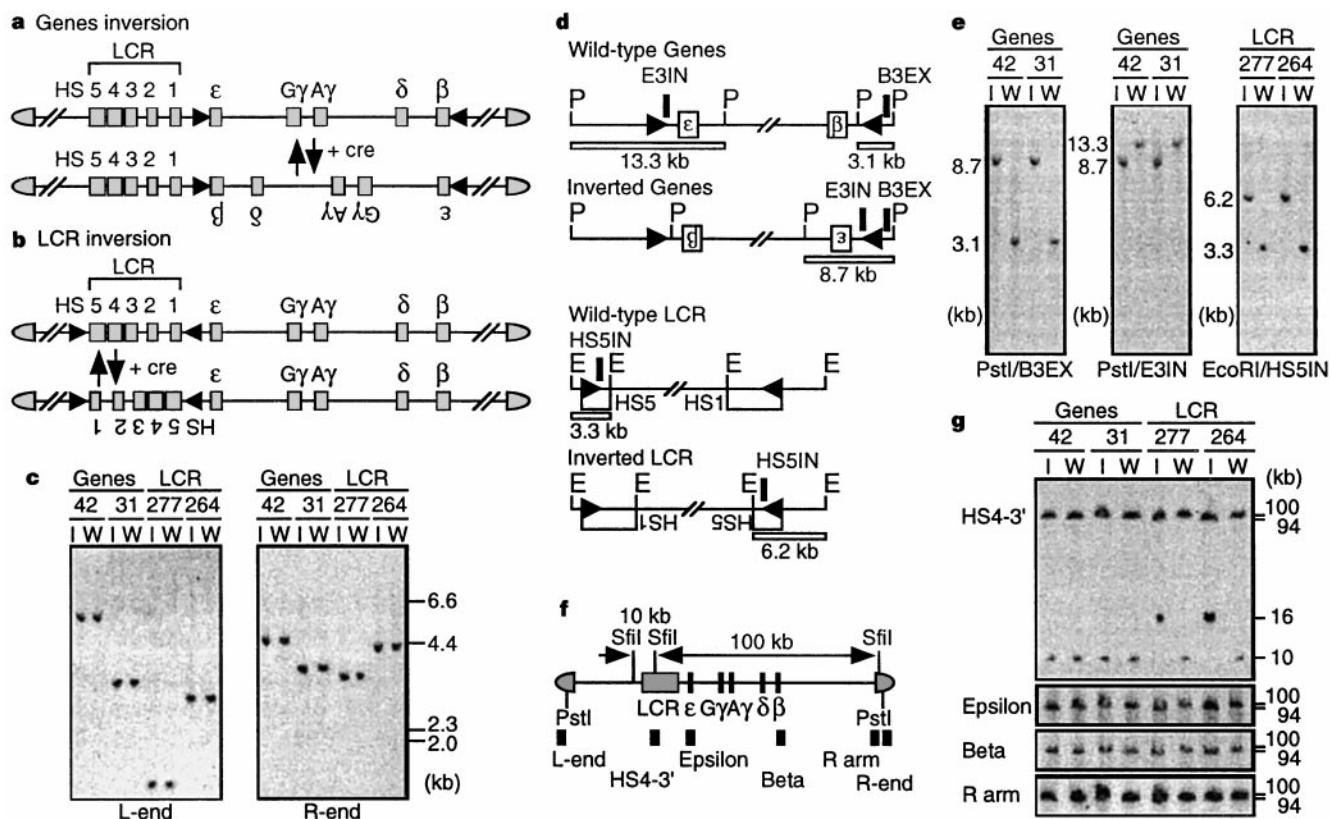


Figure 1 Structural analysis of the human β -globin YACs. **a**, Structures of wild-type and gene-inverted human β -globin YACs. *LoxP* sites (arrowheads) were introduced into the human β -globin YAC in inverted orientation. Cre-*loxP* recombination was used to create the inverted locus. HS, hypersensitive site. **b**, Structures of wild-type and LCR-inverted human β -globin YACs. **c**, *PstI*-digested tail DNA was probed with L-end (derived from the left vector arm of the YAC) and R-end (from the right YAC arm) probes (**f**). Genes, wild-type or gene-inverted YAC; LCR, wild-type or LCR-inverted YAC; 42, 31, 277, 264, transgenic mouse lines; I, inverted genes or LCR; W, wild-type genes or LCR. **d**, Detailed structures around the *loxP* sites in the wild-type and INV transgenes. The positions of the *loxP* sites (arrowheads), genes (labelled boxes), restriction enzyme sites (P, *PstI*; E, *EcoRI*)

and probes (black rectangles) are shown proportionally. Expected fragments produced by restriction enzymes are shown as open bars together with their sizes. **e**, *PstI*- or *EcoRI*-digested tail DNA was hybridized with the probes shown in **d**. **f**, Structure of the wild-type human β -globin YAC (showing positions of *SfiI* sites). The whole β -globin locus is contained within two *SfiI* fragments (of lengths 10 and 100 kb in the wild-type and gene-inverted loci and 16 and 94 kb in the LCR-inverted locus). The positions of the probes used for PFGE Southern blots are shown by black rectangles. **g**, DNA from thymus cells was digested with *SfiI* in agarose plugs, separated by PFGE, and hybridized separately to probes (indicated on the left of each panel) from the β -globin locus or from the right YAC vector arm.

locus after *in vivo* recombination, we used the E3IN probe (which corresponds to an area 5' of the ϵ -globin gene; Fig. 1d) to detect an 8.7-kb *PstI* fragment in the inverted (INV) locus rather than the 13.3-kb *PstI* fragment in the wild-type locus (Fig. 1e). The same 8.7-kb restriction fragment also hybridized to a second probe, B3EX, corresponding to the 3' end of the β -globin gene. This probe detected a 3.1-kb fragment in the wild-type locus (Fig. 1e). Similar analyses confirmed the inversion of the LCR by Cre-mediated recombination (Fig. 1d, e). As the bands detected in the end-fragment assay (Fig. 1c) were the same size in both the wild-type and the INV loci, the mutant YAC transgenes were unchanged in chromosomal position after Cre-mediated recombination.

The whole β -globin locus is included within two *SfiI* restriction-enzyme fragments in modified YAC A201F4.3 (Fig. 1f), and hybridization data (Fig. 1g) showed that all of the transgenic animals contained at least one intact human β -globin YAC; no extra bands were detected. Thus Cre-mediated site-specific recombination *in vivo* reversed the orientation of the genes, or the LCR, without otherwise altering the structure of the transgenes. Therefore this strategy eliminates the possibility of integration-dependent differences between the wild-type and mutant transgene loci, and thus any expression differences must result solely from the orientation of the mutant versus wild-type genes or LCR.

For expression analysis, we isolated RNA from the spleens of two adult anaemic mice from each transgenic line and assayed this RNA by semiquantitative reverse transcription with PCR (RT-PCR)⁹. The relative abundance of β -globin messenger RNA was similar in both the INV and the wild-type transgenic mice (Fig. 2a), confirming the results of earlier experiments that indicated that the distance or relative order of the β -globin gene with respect to the LCR is not important for its expression in the adult. Expression of the β -globin gene in fetal liver (of two independent litters) at 14.5 days post-coitum was also not significantly altered in the INV mutants (Fig. 2c). However, expression of the γ -globin genes was not detectable in the INV transgenic mice (Fig. 2b), indicating that the presence of a more strongly competing gene (that is, β -globin) between the LCR and the γ -globin genes may interfere with γ -globin-gene activation in fetal liver definitive erythroid cells, in agreement with previous conclusions⁵.

The most severe effect of inverting the genes was observed in embryonic yolk-sac erythropoiesis at 9.5 days post-coitum. In INV transgenic mice, the β -globin gene (now closest to the LCR) was expressed, whereas the ϵ -globin gene was completely silent (Fig. 2d, e). A possible competitive role for the ϵ -globin gene had not been demonstrated previously, but these data indicate that the ϵ -globin gene in the yolk sac stage may be, like the γ - and β -globin genes in

adult erythroid cells^{3,15}, regulated by competition for the activity of the LCR¹⁶. Alternatively, it is possible that, in primitive erythroid-lineage cells, the LCR, or a redundant chromatin-modulatory activity¹⁷, is simply incapable of opening chromatin as far away as the position of the most distant gene. In the INV embryonic yolk sac, the expression of the γ -globin genes was also reduced by roughly half in comparison with the wild-type genes (Fig. 2f), indicating that the β -globin gene may compete more effectively with the γ -globin genes for LCR activity than does the ϵ -globin gene in the yolk-sac environment⁵. We then determined whether inversion of the genes relative to the LCR would affect the relative

expression levels of the two very similar, but distinguishable, γ -globin genes (Fig. 2g). In the wild-type locus, ~70% of γ -globin-gene expression in the yolk sac was contributed by the $G\gamma$ -globin gene, which lies closer to the LCR. In contrast, when the $A\gamma$ -globin gene was closer to the LCR (in the INV locus), more than 60% of total γ -globin-gene transcription was contributed by the $A\gamma$ gene (Fig. 2h, i). Thus inversion of the genes results in stimulation of transcription of one gene, with simultaneous and reciprocal depression of transcription of another, the hallmark of competition between regulatory sequences^{6,7}.

In the yolk-sac stage of transgenic mice, the human adult β -globin gene is active when it is close to the LCR³⁻⁵, and therefore all of the human β -type-globin promoters (ϵ , $G\gamma$, $A\gamma$ and β) are potentially active during the embryonic stage in mice. However, in the wild-type human β -globin locus, only the ϵ - and γ -globin genes are expressed in the yolk sac, and it is evident from these and other studies that proximity to the LCR is a critical determinant for gene activity during this stage of erythropoiesis. But the fact that, in the wild-type locus, the γ -globin genes are expressed at higher levels than the ϵ -globin gene shows that distance from the LCR is a secondary determinant of globin-gene transcriptional activity.

We next studied the effects of LCR inversion. We found that, in LCR-inverted transgenic mice, transcription of all of the β -like-globin genes, at all developmental stages, was severely diminished (Fig. 3).

There are several possible explanations for orientation-dependent activity of the LCR. First, hypersensitive site 5 may have intrinsic enhancer-blocking activity¹⁸, although this has not yet been confirmed in animals¹⁹. It is, therefore, conceivable that by inverting the LCR, thus placing hypersensitive site 5 between the globin genes and the rest of the LCR, the activating elements (hypersensitive sites 2-4)²⁰ are no longer able to potentiate globin-gene transcription. Our observation that hypersensitive sites are formed in the LCR-inverted mice (data not shown) is consistent with this possibility.

Second, hypersensitive site 5 contains sequences homologous to matrix-attachment regions (MARs)²¹. If this site functions as a 5' 'boundary' element in the wild-type β -globin locus, the inverted LCR might no longer be able to overcome the repressive effects of neighbouring chromatin. In agreement with this 'hypersensitive-

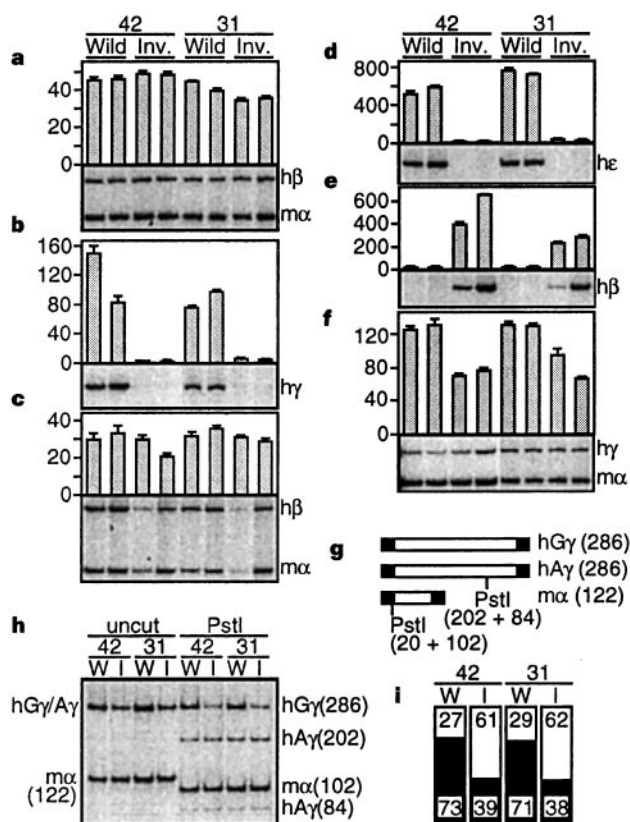


Figure 2 Expression of human β -like-globin genes in gene-inverted YAC transgenic mice. **a**, Semiquantitative RT-PCR was performed with two sets of primers for mouse α -globin (12 cycles) and human adult β -globin (12 cycles) transcripts. Representative data are shown beneath each panel. Band intensities were quantified by phosphorimaging and normalized expression levels (see Methods) are shown as histograms. **b**, **c**, Fetal RNA was analysed for the expression of human γ -globin (**b**, 18 cycles) or mouse α -globin and human β -globin (**c**, 12 cycles) genes. **d**–**f**, RNA from two independent litters was analysed for the expression of human ϵ -globin (**d**, 18 cycles), human β -globin (**e**, 18 cycles), or mouse α -globin and human γ -globin (**f**, 12 cycles) genes in the embryonic (9.5 days post-coitum) yolk sac. **g**, RT-PCR analysis of human $G\gamma$: $A\gamma$ transcript ratios in yolk sacs of transgenic mice. RNA samples were amplified with primer sets specific for human $G\gamma$ and $A\gamma$ and mouse α -globin genes. The positions of the primers (black boxes), the total lengths of each RT-PCR product (open box; sizes in base pairs are in parentheses), and the positions of the *PstI* restriction sites, together with the fragment sizes produced, are indicated. **h**, PCR products were digested with *PstI* and separated on a polyacrylamide gel. A *PstI* site present in the $A\gamma$, but not in the $G\gamma$, gene enabled separation and quantification of products derived from the individual γ -globin-gene transcripts. To control for *PstI*-mediated digestion, a *PstI* site was introduced into the α -globin primer. **i**, The contributions of the $G\gamma$ and $A\gamma$ genes relative to total γ -globin synthesis (arbitrarily set at 100%) are plotted as filled and open bars, respectively. Average values obtained from two independent experiments are shown within each of the bars. W, wild-type genes; I, inverted genes.

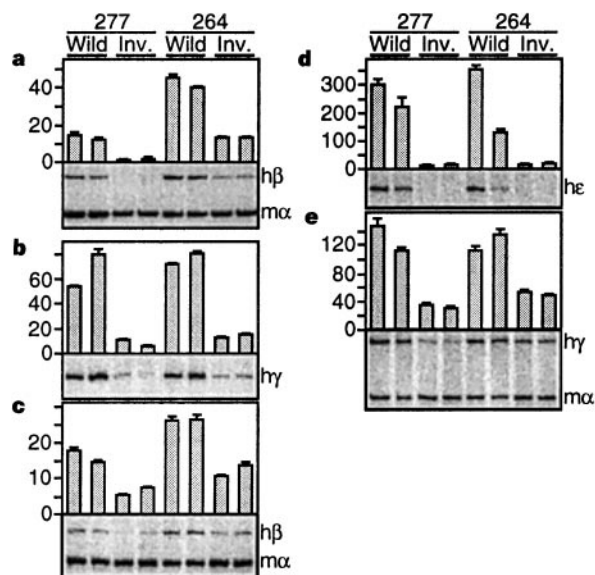


Figure 3 Expression of human β -like-globin genes in LCR-inverted YAC transgenic mice. **a**, β -globin-gene expression in adult spleens. **b**, γ - and **c**, β -globin-gene expression in fetal liver at 14.5 days post-coitum. **d**, ϵ - and **e**, γ -globin-gene expression in embryonic yolk sac at 9.5 days post-coitum. W, wild-type LCR; I, inverted LCR.

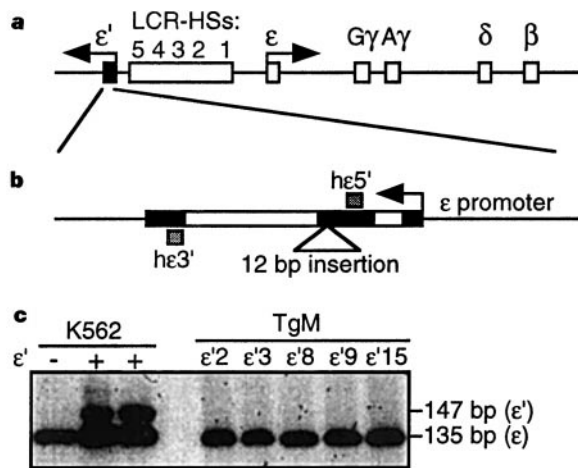


Figure 4 ϵ -globin is not expressed when positioned 5' of the LCR. **a**, ϵ' -globin gene (black rectangle) was inserted 0.6 kb 5' of the *Hind*III site in hypersensitive site 5 (HS5) in the opposite transcriptional orientation to the wild-type ϵ -globin gene. **b**, The human ϵ -globin gene was generated as a *Sph*I-*Hind*III fragment (3.8 kb) and a 12-base-pair (bp) oligonucleotide was then inserted at the *Bam*HI site (nucleotide position 19,959) in exon 2. The positions of exons (filled boxes), introns (open boxes) and the primers used⁹ for RT-PCR analysis of ϵ - and ϵ' -globin RNA (shaded boxes) are shown. **c**, Left three lanes, the ϵ' gene was linked to human HS2 and transfected into human embryonic erythroid K562 cells. RNA was isolated from non-transfected (–) or transfected (+) cells and analysed for both endogenous ϵ -globin (135 bp) and transfected ϵ' -globin (147 bp) transcripts by RT-PCR. Right five lanes, at the same time, RNA from yolk sacs of five independent lines (2, 3, 8, 9 and 15) of human β -globin/ ϵ' -globin YAC transgenic mice at 9.5 days post-coitum was analysed for the accumulation of ϵ -globin and ϵ' -globin mRNA. These lines were analysed for transgene integrity and copy number as described in the legend to Fig. 1.

site 5 insulator' hypothesis, we found that a marked ϵ -globin gene, when placed 5' of the LCR, was not expressed in YAC transgenic mice (Fig. 4). However, in another example, the β -globin gene was, in fact, transcribed from a position 5' of the LCR¹⁹. The number of differences between our strategy and that used in ref. 19 makes the origin of the discrepancy in our results uncertain.

Third, LCR inversion might interfere with the formation of a higher-order chromatin architecture, by separating functionally vital interactions between known elements located between the *loxP* sites from uncharacterized ones lying outside the sites.

In conclusion, the ϵ -globin gene, when relocated to the 3' end of the β -globin locus, is transcriptionally silent at all developmental stages, showing that its normal proximity to the LCR is required for yolk-sac expression of the ϵ -globin gene. The β -globin gene, when in the LCR-proximal position, is transcribed at all stages, showing that competition within the wild-type locus normally prevents premature β -globin-gene expression during the yolk-sac developmental stage. When the LCR is inverted *in vivo*, transcription of all of the β -type globin genes is severely impaired at all stages, indicating that the β -globin LCR acts in an orientation-dependent manner, and that it therefore does not fit the simplest definition of an enhancer. Thus these data provide, to our knowledge, the first *in vivo* demonstration that the LCR is orientation dependent, the mechanistic basis for which is being investigated at present. □

Methods

YAC mutagenesis. *LoxP* sites were inserted into the human β -globin YAC A201F4.3 (150 kb in length) at the following positions: 5' of hypersensitive site 5 (at position 199; ref. 21) and 3' of hypersensitive site 1 (position 13,769; <http://globin.cse.psu.edu>); or between the LCR and the ϵ -globin gene (at position 17,482) and 3' of the β -globin gene enhancer (at position 65,421).

The targeting sequences were inserted into yeast integrative plasmid pRS306 to generate the appropriate yeast vectors used to insert *loxP* sites into the YAC. Yeast methods used have been described previously⁹.

Transgenic mice. YAC DNA was isolated from pulsed-field gels according to standard protocols, with minor modifications⁹, and injected into fertilized mouse eggs (CD1 eggs; Charles River Breeding Laboratory)²². For *in vivo* recombination, the plasmid pBS185 (which expresses the recombinase Cre under the control of the cytomegalovirus promoter²³) was injected into pronuclei and the eggs were transferred to foster mothers. Southern blot analysis was performed using DNA from transgenic tails or thymi. The preparation of thymus DNA in agarose plugs for pulsed-field gel electrophoresis has been described²⁴.

Semiquantitative RT-PCR. Total RNA was isolated from erythroid tissues and used for RT-PCR as described⁹. The following PCR primers were used: human ϵ -globin, 5'-CAATCACTAGCAAGCTCTCAGG-3' (sense), 5'-GGGCTTG-AGGTTGTCCATGTTT-3' (antisense); human γ -globin, 5'-GATGCCAT-AAAGCACCTGGATG-3' (sense), 5'-TTGCAGAATAAAGCCTATCCTTGA-3' (antisense); human β -globin, 5'-AACTGTGTTCACTAGCAACCTCAA-3' (sense), 5'-GAGTGGACAGATCCCCAAAGGA-3' (antisense); mouse α -globin, 5'-GATTCTGACAGACTCTGCAGGAAAC-3' (sense); the *Pst*I site is underlined, 5'-CCTTTCCAGGGCTTCAGCTCCATAT-3' (antisense). Because of the difficulty in comparing low-level expression (for example, of the γ -globin gene in the fetal liver) with abundant expression (for example, of the β - or α -globin genes in the fetal liver), genes were not always analysed at the same PCR cycle number. When analysing more than one gene per reaction, a master mix containing all components (including complementary DNA template) except the primers was first prepared. Excess primer sets specific for the genes were then added to individual aliquots. Thus the same amount of cDNA template was present in each tube and the signal could be normalized for different PCR cycle numbers. PCR was performed for the cycle numbers shown in the figures and the ratio of human β -type-globin gene signal to that of endogenous mouse α -globin gene at 12 cycles (baseline arbitrarily set as 100) was calculated. At least three independent experiments for each sample were performed to eliminate pipetting error and the mean $\pm$ s.d. is shown.

Received 20 November 1998; accepted 11 February 1999.

1. Stamatoyannopoulos, G. & Nienhuis, A. W. In *The Molecular Basis of Blood Diseases* 2nd edn (eds Stamatoyannopoulos, G., Nienhuis, A. W., Majerus, P. & Varmus, H.) 107–155 (Saunders, New York, 1994).
2. Hardison, R. *et al.* Locus control regions of mammalian β -globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* **205**, 73–94 (1997).
3. Hanscombe, O. *et al.* Importance of globin gene order for correct developmental expression. *Genes Dev.* **5**, 1387–1394 (1991).
4. Peterson, K. R. & Stamatoyannopoulos, G. Role of gene order in developmental control of human γ - and β -globin gene expression. *Mol. Cell. Biol.* **13**, 4836–4843 (1993).
5. Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. & Grosfeld, F. The effect of distance on long-range chromatin interactions. *Mol. Cell* **1**, 131–139 (1997).
6. Choi, O.-R. & Engel, J. D. Developmental regulation of β -globin gene switching. *Cell* **55**, 17–26 (1988).
7. Foley, K. P. & Engel, J. D. Individual stage selector element mutations lead to reciprocal changes in β - vs. ϵ -globin gene transcription: genetic confirmation of promoter competition during globin gene switching. *Genes Dev.* **6**, 730–744 (1992).
8. Martin, D. I. K., Fiering, S. & Groudine, M. Regulation of β -globin gene expression: straightening out the locus. *Curr. Opin. Genet. Dev.* **6**, 488–495 (1996).
9. Bungert, J. *et al.* Synergistic regulation of human β -globin gene switching by locus control region elements HS3 and HS4. *Genes Dev.* **9**, 3083–3096 (1995).
10. Milot, E. *et al.* Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**, 105–114 (1996).
11. Sauer, B. & Henderson, N. Cre-stimulated recombination at *loxP*-containing DNA sequences placed into the mammalian genome. *Nucleic Acids Res.* **17**, 147–161 (1989).
12. Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
13. Gaensler, K. M., Kitamura, M. & Kan, Y. W. Germ-line transmission and developmental regulation of a 150-kb yeast artificial chromosome containing the human β -globin locus in transgenic mice. *Proc. Natl Acad. Sci. USA* **90**, 11381–11385 (1993).
14. Liu, Q., Bungert, J. & Engel, J. D. Mutation of gene-proximal regulatory elements disrupts human ϵ -, γ -, and β -globin expression in yeast artificial chromosome transgenic mice. *Proc. Natl Acad. Sci. USA* **94**, 169–174 (1997).
15. Enver, T. *et al.* Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* **344**, 309–313 (1990).
16. Wijgerde, M., Grosfeld, F. & Fraser, P. Transcription complex stability and chromatin dynamics *in vivo*. *Nature* **377**, 209–213 (1995).
17. Reik, A. *et al.* The locus control region is necessary for gene expression in the human β -globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol. Cell. Biol.* **18**, 5992–6000 (1998).
18. Li, Q. & Stamatoyannopoulos, G. Hypersensitive site 5 of the human β locus control region functions as a chromatin insulator. *Blood* **84**, 1399–1401 (1994).
19. Zafarana, G., Raguz, S., Pruzina, S., Grosfeld, F. & Meijer, D. In *Molecular Biology of Hemoglobin Switching* (ed. Stamatoyannopoulos, G.) 39–44 (Intercept, Andover, 1995).

20. Fraser, P., Pruzina, S., Antoniou, M. & Grosfeld, F. Each hypersensitive site of the human β -globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev.* **7**, 106–113 (1993).
21. Yu, J., Bock, J. H., Slightom, J. L. & Villeponteau, B. A 5' β -globin matrix-attachment region and the polyoma enhancer together confer position-independent transcription. *Gene* **139**, 139–145 (1994).
22. Hogan, B., Costantini, F. & Lacy, E. *Manipulating the Mouse Embryo* (Cold Spring Harbor Press, Cold Spring Harbor, 1986).
23. Sauer, B. & Henderson, N. Targeted insertion of exogenous DNA into the eucaryotic genome by the cre recombinase. *New Biol.* **2**, 441–449 (1990).
24. Liu, Q., Tanimoto, K., Bungert, J. & Engel, J. D. The A γ -globin 3' element provides no unique function(s) for human β -globin locus gene regulation. *Proc. Natl Acad. Sci. USA* **95**, 9944–9949 (1998).

Acknowledgements. We thank Y. Tanimoto, J. Fan and R. Arora for help with this work. We also acknowledge the support of the Japan Society for the Promotion of Science and the Naito Foundation (K.T.), the Cooley's Anemia Foundation (K.T. and Q.L.) and the American Heart Association (J.B.) for fellowship support, and the Robert H. Lurie Comprehensive Cancer Center and the NIH for research support (J.B. and J.D.E.)

Correspondence and requests for materials should be addressed to J.D.E. (e-mail: d-engel@nwu.edu).

Chaperone-like activity of the AAA domain of the yeast Yme1 AAA protease

Klaus Leonhard, Alexandra Stiegler, Walter Neupert & Thomas Langer

Institut für Physiologische Chemie der Universität München, Goethestrasse 33, 80336 München, Germany

The AAA domain, a conserved Walker-type ATPase module, is a feature of members of the AAA family of proteins^{1,2}, which are involved in many cellular processes, including vesicular transport^{3–7}, organelle biogenesis⁸, microtubule rearrangement⁹ and protein degradation^{10–12}. The function of the AAA domain, however, has not been explained. Membrane-anchored AAA proteases of prokaryotic and eukaryotic cells comprise a subfamily of AAA proteins^{13–15} that have metal-dependent peptidase activity and mediate the degradation of non-assembled membrane proteins. Inactivation of an orthologue of this protease family in humans causes neurodegeneration in hereditary spastic paraplegia¹⁶. Here we investigate the AAA domain of the yeast protein Yme1, a subunit of the *i*-AAA protease located in the inner membrane of mitochondria^{17,18}. We show that Yme1 senses the folding state of solvent-exposed domains and specifically degrades unfolded membrane proteins. Substrate recognition and binding are mediated by the amino-terminal region of the AAA domain. The purified AAA domain of Yme1 binds unfolded polypeptides and suppresses their aggregation. Our results indicate that the AAA domain of Yme1 has a chaperone-like activity and suggest that the AAA domains of other AAA proteins may have a similar function.

Two model proteins were employed to examine the mechanism of substrate recognition by mitochondrial AAA proteases: the amino-terminal parts of these proteins consist of the 161 N-terminal amino acid residues of the mitochondrial inner membrane protein Yta10, including its mitochondrial targeting sequence and its first membrane-spanning segment¹⁹; the carboxy-terminal part comprises the wild-type form of mouse dihydrofolate reductase (DHFR) or a loosely folded, mutant variant of it²⁰ (Yta10(161)-DHFR^{WT}, Yta10(161)-DHFR^{MUT}). These fusion proteins were synthesized in a cell-free system and imported into isolated mitochondria. Both proteins were inserted into the mitochondrial inner membrane. We found that their DHFR domains were exposed to the intermembrane space (data not shown).

When mitochondria were further incubated at 25 °C after completion of import, loosely folded Yta10(161)-DHFR^{MUT}, but not the fusion protein containing wild-type DHFR, was rapidly degraded (Fig. 1a). Proteolysis was impaired after mitochondria were

depleted of ATP or in the presence of the non-hydrolysable ATP analogue β - γ -imidoadenosine 5'-phosphate (AMP-PNP) (Fig. 1b). Yta10(161)-DHFR^{MUT} accumulated in mitochondria lacking Yme1 or in mitochondria harbouring proteolytically inactive variants of Yme1 with point mutations either in the proteolytic site (*yme1*^{E541Q})^{17,18} or in the conserved Walker A motif of the AAA domain (*yme1*^{K327R})¹⁸ (Fig. 1b). Thus, proteolysis depends on ATP hydrolysis and is mediated by Yme1, which constitutes the *i*-AAA protease. This protease is active at the outer surface of the inner membrane¹⁷ and has previously been shown to degrade non-assembled Cox2 in the inner membrane^{18,21,22}. In contrast, Yta10(161)-DHFR^{WT} was degraded by Yme1 only upon incubation of mitochondria at high temperature (Fig. 1c). This results from thermal unfolding of the DHFR domain because the folic acid analogue methotrexate, which is known to stabilize the native conformation of DHFR, blocked proteolysis²³ (Fig. 1c). We conclude from these experiments that unfolding of the solvent-exposed domain of an integral membrane protein can trigger its proteolysis by Yme1.

The association of both model proteins with Yme1 was assessed by immunoprecipitation with a polyclonal antiserum directed against the N-terminus of mature Yme1¹⁷. Newly imported Yta10(161)-DHFR^{MUT} was co-precipitated with proteolytically inactive Yme1 variants containing a point mutation either in the consensus binding site for ATP or in the proteolytic site (Fig. 2a). In contrast, the fusion protein with the wild-type form of DHFR was

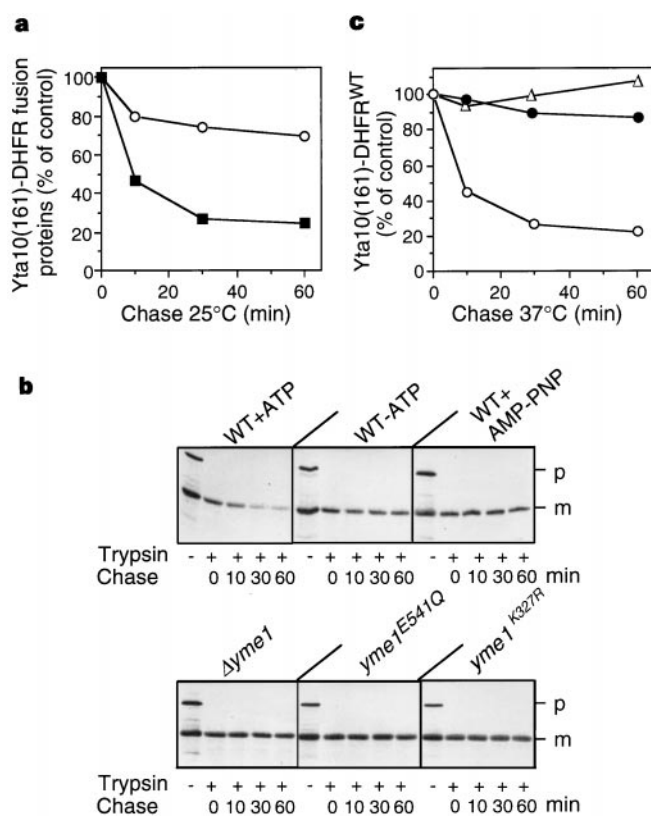


Figure 1 Unfolding of a solvent-exposed domain can trigger the degradation of a membrane protein by Yme1. **a**, Stability of newly imported Yta10(161)-DHFR^{WT} (open circles) and Yta10(161)-DHFR^{MUT} (filled squares) in wild-type mitochondria at 25 °C. **b**, Stability of newly imported Yta10(161)-DHFR^{MUT} in wild type (WT), Δ yme1, yme1^{E541Q} and yme1^{K327R} mitochondria at 25 °C. When indicated, ATP-levels were decreased by replacing the ATP-regenerating system by apyrase (5 U ml⁻¹, -ATP) or by apyrase and AMP-PNP (4 mM). p, precursor form; m, mature form. **c**, Stability of Yta10(161)-DHFR^{WT} in wild-type (circles) and Δ yme1 (triangles) mitochondria at 37 °C. Methotrexate (3 μ M) was added (filled circles) after import.

not found in association with the protease (Fig. 2a). These results indicate that the folding state of a solvent-exposed domain of a membrane protein is recognized by Yme1, which binds only if this domain is unfolded. Binding of Yta10(161)-DHFR^{MUT} to proteolytically active Yme1 was hardly detectable, presumably owing to its rapid proteolysis (Fig. 2a). Consistently, inhibition of proteolysis by the depletion of ATP in mitochondria resulted in a slightly increased efficiency of coprecipitation (Fig. 2b, upper panel). Yta10(161)-DHFR^{MUT}, which was bound to wild-type Yme1 under these conditions, remained associated with the protease upon further incubation in the absence of ATP (Fig. 2b, upper panel). Similarly, we observed binding but not release of Yta10(161)-DHFR^{MUT} from proteolytically inactive Yme1^{E541Q} or Yme1^{K327R} (Fig. 2b, lower panel). Mutations analogous to *yme1*^{K327R} in the ATP-binding site of related AAA proteins were reported to abolish nucleotide binding^{7,24}. We therefore conclude that the association of Yme1 with an unfolded polypeptide does not depend on the presence of ATP, whereas the degradation of bound substrates does require ATP hydrolysis.

Two conserved domains can be distinguished in the C-terminal part of Yme1 (Fig. 2c): the AAA domain characteristic of the AAA family of ATPases^{1,12} is followed by a domain that is found only in AAA proteases and contains the proteolytic site. To examine whether the proteolytic domain of Yme1 is required for binding

of misfolded polypeptides, we constructed C-terminally truncated variants of Yme1 and expressed them in $\Delta yme1$ cells. All mutant forms of Yme1 were properly integrated in the mitochondrial inner membrane (data not shown). The binding of newly imported Yta10(161)-DHFR^{MUT} occurred with similar efficiencies to proteolytically inactive Yme1^{E541Q} and to Yme1¹⁻⁵¹⁴ lacking the proteolytic centre (Fig. 2d). The proteolytic domain is thus dispensable for binding of substrate polypeptides to Yme1. In contrast, deletion of the AAA domain of Yme1 (Yme1¹⁻²⁶⁷) strongly impaired binding of Yta10(161)-DHFR^{MUT} (Fig. 2d). Although a contribution from other domains of Yme1 cannot be excluded, the presence of the AAA domain is apparently crucial for efficient substrate binding.

Sequence alignments of AAA proteins with other Walker-type NTPases have led to the identification of regions within the AAA domain that are characteristic of AAA-like proteins but are absent from other NTPases^{1,12} (Fig. 2c). The N-terminal region of the AAA domain was found to be crucial for substrate binding. In contrast to Yme1¹⁻²⁶⁷, which completely lacks the AAA domain, Yme1¹⁻³¹³ was capable of binding newly imported Yta10(161)-DHFR^{MUT} with an efficiency similar to that of proteolytically inactive Yme1^{E541Q} (Fig. 2d).

The AAA domain of Yme1 might be involved in the interaction with other, so far unidentified, subunits of the *i*-AAA protease needed for substrate binding. To exclude this possibility, we expressed the whole intermembrane space domain of Yme1 (residues 250–747; IMS) and the AAA domain of Yme1 (residues 250–525), both containing a hexahistidine peptide fused to the C-terminus, in *Escherichia coli* and purified them to homogeneity

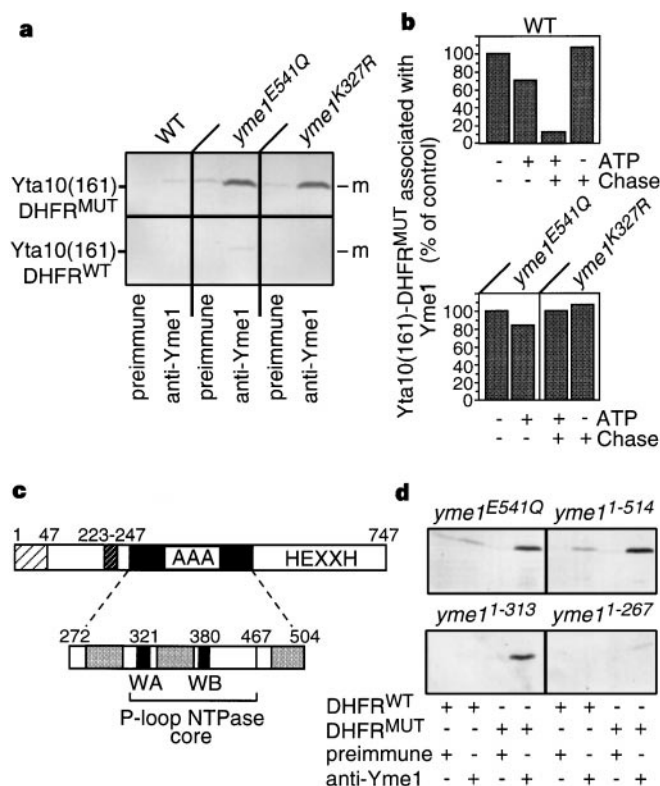


Figure 2 Interaction of Yta10(161)-DHFR^{WT} and Yta10(161)-DHFR^{MUT} with Yme1. **a**, Co-immunoprecipitation experiments with preimmune and Yme1-specific antisera after the import of both fusion proteins into wild-type (WT), *yme1*^{E541Q} and *yme1*^{K327R} mitochondria. **b**, Co-immunoprecipitation of newly imported Yta10(161)-DHFR^{MUT} with Yme1 after a further incubation of mitochondria for 30 min at 25°C in the presence of different ATP levels (chase). Yta10(161)-DHFR^{MUT} that associated with Yme1 in the absence of ATP before the chase was set to 100%. **c**, Domain structure of Yme1. 1–47, mitochondrial targeting sequence; 223–247, transmembrane segment; 272–504, AAA domain containing Walker motifs A and B (WA, WB) (AAA typical regions are shown in grey boxes); 504–747, proteolytic domain with metal-binding site HEXXH. **d**, Co-immunoprecipitation of Yta10(161)-DHFR^{WT} and Yta10(161)-DHFR^{MUT} with proteolytically inactive Yme1^{E541Q} and C-terminally truncated variants of Yme1.

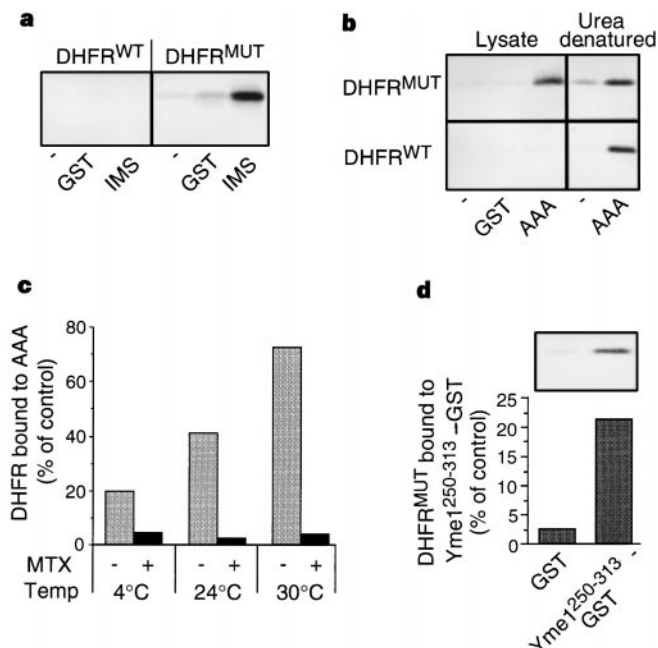


Figure 3 Binding of unfolded polypeptides to the AAA domain *in vitro*. Ni-NTA precipitation of newly synthesized DHFR^{WT} and DHFR^{MUT} with His-tagged IMS (**a**) or AAA (**b**) domains of Yme1, or with his-tagged GST. Radiolabelled substrate polypeptides were added to the isolated domains either directly (lysate) or after denaturation with urea. **c**, Temperature dependence of binding of DHFR^{WT} to the AAA domain. DHFR^{WT} was incubated for 30 min in buffer A containing AAA domain (18 μ M) in the presence or absence of methotrexate (3 μ M; MTX) at the temperatures indicated. Binding of an equal amount of unfolded Yta10(161)-DHFR^{MUT} was set to 100%. **d**, Interaction of DHFR^{MUT} with Yme1²⁵⁰⁻³¹³-GST. Bound material was precipitated with glutathione beads and quantified by phosphorimaging analysis. Binding of DHFR^{MUT} to IMS was set to 100%.

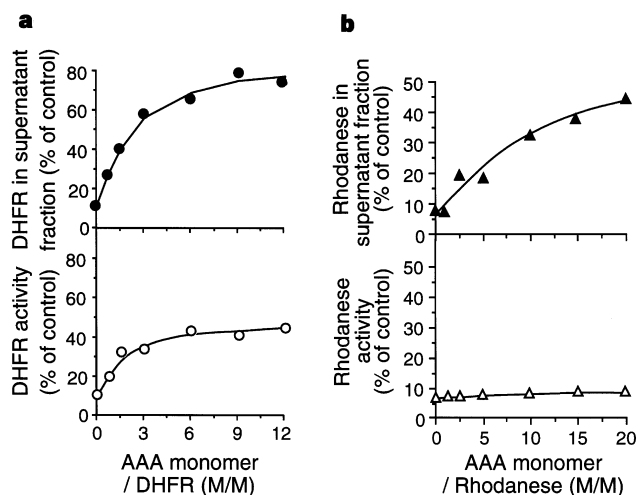


Figure 4 Chaperone-like activity of the AAA domain *in vitro*. **a**, The AAA domain suppresses DHFR aggregation and mediates its reactivation. **b**, Suppression of rhodanese aggregation by the AAA domain. DHFR (**a**) or rhodanese (**b**) were denatured in 7 M urea, 30 mM Tris-HCl pH 7.4 and 50 mM dithiothreitol or in 7 M guanidinium chloride, 30 mM Tris-HCl pH 7.4 and 200 mM 2-mercaptoethanol, respectively, and diluted 100-fold (2.3 μ M) at 4°C with buffer A containing the AAA domain at various concentrations. After centrifugation for 20 min at 30,000g, the supernatant fractions were analysed by SDS-PAGE and immunoblotting. The total amount of DHFR or rhodanese in the assay was set to 100%. Enzymatic activities of DHFR and rhodanese in the supernatant fraction (open symbols) were determined as described^{28,29}. The activity of native enzymes was set to 100%.

(data not shown). DHFR and its loosely folded mutant variant DHFR^{MUT} were synthesized in a cell-free translation system in the presence of [³⁵S]methionine and added to the purified proteins. Substrate binding was assessed by precipitation with Ni-nitrilotriacetate (NTA)-agarose beads. For control, DHFR and DHFR^{MUT} were incubated in the absence of the domains or in the presence of glutathione-S-transferase (GST) containing a C-terminal hexahistidine peptide; we then examined substrate binding using Ni-NTA beads. DHFR^{MUT} was associated with the AAA domain and the IMS domain of Yme1 but not with GST (Fig. 3a, b). In contrast, wild-type DHFR was hardly detectable in the precipitate (Fig. 3a, b). Both DHFR and DHFR^{MUT} were bound to the AAA domain when precubated with a high concentration of urea (Fig. 3b). Furthermore, thermally denatured wild-type DHFR was found in association with the AAA domain. Binding to the AAA domain was hardly detectable at 4°C but increased with increasing temperature (Fig. 3c). The ability of methotrexate to prevent binding at all temperatures tested is consistent with a requirement of the unfolded DHFR for interaction with the AAA domain (Fig. 3c).

To examine whether N-terminal amino acid residues of the AAA domain are sufficient for substrate binding, we expressed a GST fusion protein containing residues 250–313 of Yme1 in *E. coli* and purified it to homogeneity (data not shown). The association of newly synthesized DHFR^{MUT} with Yme1^{250–313}-GST was assessed by using a glutathione resin. DHFR^{MUT} was specifically precipitated with glutathione beads in the presence of Yme1^{250–313}-GST (Fig. 3d). The binding efficiency, however, was ~5-fold lower in these experiments when compared with the full-length IMS domain. Nevertheless, this observation, and the co-immunoprecipitation experiments described above, assign an important role for substrate binding to the N-terminal region of the AAA domain.

These results demonstrate that the AAA domain of Yme1 specifically binds unfolded polypeptides, a property reminiscent of molecular chaperones. To provide further evidence for a chaper-

one-like activity, we examined the ability of the AAA domain to suppress aggregation and promote the refolding of denatured polypeptides. Purified DHFR was denatured with urea and diluted into buffer containing the AAA domain or, as a control, GST at various concentrations. DHFR forms insoluble aggregates under these conditions. Aggregation was strongly decreased in the presence of the AAA domain, whereas GST did not affect the solubility of DHFR (Fig. 4a and data not shown). DHFR activity was detected in the soluble fractions, indicating that the AAA domain mediated DHFR reactivation (Fig. 4a). In similar experiments, the AAA domain of Yme1 also suppressed the aggregation of rhodanese in a concentration-dependent manner (Fig. 4b). However, we did not observe reactivation of urea-denatured rhodanese (Fig. 4b). Thus, the AAA domain of Yme1 prevents the aggregation of structurally unrelated polypeptides and, presumably depending on their folding properties *in vitro*, can mediate their refolding. The ability of the AAA domain to mediate the reactivation of urea-denatured DHFR *in vitro* substantiates its chaperone-like activity. *In vivo* evidence for non-proteolytic functions of Yme1, however, is lacking so far. It is therefore conceivable that the chaperone-like function of the AAA domain is restricted to proteolytic substrates of Yme1.

We observed efficient suppression of aggregation in these experiments only if the AAA domain was present in a molar excess over unfolded polypeptides (Fig. 4). The apparently low substrate affinity of the isolated AAA domain might be a consequence of its oligomeric state *in vitro*: in contrast to Yme1, which forms a multi-subunit complex in the membrane¹⁷, gel-filtration analysis revealed a native relative molecular mass (M_r) of the isolated AAA domain of ~100,000, suggesting a trimeric state (data not shown). Cooperative interactions between subunits might increase the affinity of assembled Yme1 *in vivo*.

Our results demonstrate chaperone-like properties of the AAA domain of Yme1 and thereby provide a rationale for the specific recognition of non-native polypeptides by AAA proteases. We propose that the hydrolysis of ATP triggers the release of associated substrates from the AAA domain, allowing, perhaps after rebinding, their degradation by the proteolytic domain. ATP-dependent conformational changes of Yme1 might concomitantly induce the unfolding of bound polypeptides and regulate the accessibility of the proteolytic sites. The combination of an ATPase domain that has chaperone-like activity with a proteolytic domain is reminiscent of other energy-dependent proteases that harbour these activities on different subunits^{25,26}. The regulatory subunits of the hetero-oligomeric Clp proteases exert ATPase and chaperone activity and present substrate polypeptides to proteolytic subunits for degradation. Notably, similar tertiary structures for the ATPase domains of AAA, Lon and Clp proteins have been predicted on the basis of sequence comparisons¹².

The AAA domain of other members of the AAA family might have a function similar to that of Yme1. Many AAA proteins seem to be involved in the assembly or dissociation of protein complexes. NSF (N-ethylmaleimide-sensitive factor) harbours two AAA domains; it has been demonstrated to induce conformational changes in so-called SNARE proteins (required for membrane fusion) and thereby to trigger the disassembly of protein complexes and membrane fusion^{3,5,6}. Similarly, AAA proteins present in the regulatory complex of the 26S proteasome have been implicated in the binding and unfolding of substrate polypeptides and in their translocation to the proteolytic sites¹¹. Our results suggest that AAA domains have a central role in these processes by exerting chaperone-like properties. □

Methods

Cloning procedures. C-terminally truncated variants of YME1 were generated by polymerase chain reaction (PCR), cloned into the centromeric yeast expression plasmid YCplac111:ADH1 and expressed in $\Delta yme1$ cells¹⁷. The hybrid proteins Yta10(161)-DHFR^{WT} and Yta10(161)-DHFR^{MUT} were cloned

by PCR into pGEM4 (Promega) allowing SP6 polymerase-driven expression. For expression in *E. coli*, PCR-amplified fragments of *YME1* encoding residues 250–525 or 250–747 were cloned into pQE60 providing a C-terminal hexahistidine tag (pQE60-AAA^{YME1}, pQE60-IMS^{YME1}). The Yme1^{250–313}-GST fusion protein was constructed by cloning the corresponding PCR-amplified fragment of *YME1* into pGEX-4T-1.

Protein import and co-immunoprecipitation experiments. Yta10(161)-DHFR^{WT} and Yta10(161)-DHFR^{MUT} were synthesized in reticulocyte lysate in the presence of [³⁵S]methionine and imported into isolated mitochondria for 10 min at 25 °C as described²⁷. Non-imported preproteins were digested by the addition of trypsin (100 µg/ml). To assess the stability of newly imported proteins, mitochondria were further incubated in the presence of an ATP-regenerating system at the indicated temperature. For co-immunoprecipitation, mitochondria (0.5 mg ml⁻¹) were solubilized in 2% digitonin (w/v), 30 mM Tris-HCl pH 7.4, 150 mM potassium acetate, 4 mM magnesium acetate, 1 mM ATP, and 1 mM phenylmethylsulphonylfluoride (PMSF). After a clarifying centrifugation for 20 min at 109,000g, supernatants were incubated with preimmune or a polyclonal antiserum directed against the N terminus of Yme1. The efficiency of co-immunoprecipitation of newly imported Yta10(161)-DHFR^{MUT} was ~12% with yme1^{E541Q} and yme1^{K327R} mitochondria or ~2% with wild-type mitochondria. To examine the release of Yta10(161)-DHFR^{MUT} from Yme1, mitochondria were incubated for 30 min at 25 °C after the completion of import, followed by solubilization and immunoprecipitation. ATP was depleted by the addition of apyrase (40 U ml⁻¹, Sigma) when indicated.

Protein expression and purification. BL21 (DE3, pREP4) *E. coli* cells transformed with pQE60-AAA^{YME1}, pQE60-IMS^{YME1} or Yme1^{250–313}-pGEX-4T-1 were grown to an A₆₀₀ of 1.0 and induced with isopropyl-β-D-thiogalactoside (1 mM) for 30 min at 37 °C. Cells were lysed by sonication in 50 mM Tris-HCl pH 8.5, 0.4% Triton X-100, 50 mM potassium acetate, 25% sucrose (w/v), 5% glycerol (v/v) and 2 mM phenylmethylsulphonylfluoride (PMSF). The supernatant of a clarifying centrifugation was fractionated by using Ni-NTA-agarose or glutathione-Sepharose, respectively. The Ni-NTA resin was washed with 50 mM Tris-HCl pH 8.5, 50 mM potassium acetate, 5% glycerol (v/v), 10 mM imidazole and 1 mM PMSF. The AAA or IMS domain of Yme1 was eluted with 50 mM Tris-HCl pH 7.4, 50 mM potassium acetate, 400 mM imidazole and 1 mM PMSF. Peak fractions containing the AAA domain were desalted with a PD10 column and applied in 20 mM Tris-HCl pH 8.0 and 20 mM potassium acetate to a MonoQ column. Bound protein was eluted with a linear salt gradient. Yme1^{250–313}-GST was purified with glutathione-Sepharose in accordance with the guidelines of the manufacturer. Purified proteins were dialysed against buffer A (50 mM Tris-HCl pH 7.4, 50 mM potassium acetate and 5% glycerol (v/v)).

In vitro binding assays. DHFR^{WT} and DHFR^{MUT} were synthesized in reticulocyte lysate in the presence of [³⁵S]methionine, diluted 100-fold into buffer A containing the AAA or IMS domain (18 µM) and incubated for 30 min at 4 °C. Newly synthesized polypeptides were precipitated with ammonium sulphate and denatured in 7 M urea, 30 mM Tris-HCl pH 7.4 and 50 mM DTT (dithiothreitol) when indicated. After dilution with buffer A (2.5-fold) and a clarifying centrifugation, the supernatant was incubated with Ni-NTA agarose beads for 1 h at 4 °C followed by extensive washing with buffer A. Bound material was eluted with buffer A containing 400 mM imidazole and quantified by phosphorimaging analysis. In the presence of the IMS or AAA domain, ~5% of substrates added were recovered in the precipitate. Binding to Yme1^{250–313}-

GST was assessed in the same manner with glutathione-Sepharose beads. Buffer A was supplemented with 10 mM glutathione to elute bound proteins.

Received 21 December 1998; accepted 9 February 1999.

1. Beyer, A. Sequence analysis of the AAA protein family. *Protein Sci.* **6**, 2043–2058 (1997).
2. Patel, S. & Latterich, M. The AAA team: related ATPases with diverse functions. *Trends Cell Biol.* **8**, 65–71 (1998).
3. Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55–63 (1994).
4. Latterich, M., Fröhlich, K. U. & Schekman, R. Membrane fusion and the cell cycle: Cdc48p participates in the fusion of ER membranes. *Cell* **82**, 885–893 (1995).
5. Mayer, A., Wickner, W. & Haas, A. Sec18p (NSF)-driven release of Sec17p (alpha-SNAP) can precede docking and fusion of yeast vacuoles. *Cell* **85**, 83–94 (1996).
6. Hanson, P. I., Roth, R., Morisaki, H., Jahn, R. & Heuser, J. E. Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. *Cell* **90**, 523–535 (1997).
7. Babst, M., Wendland, B., Estepa, E. J. & Emr, S. D. The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J.* **17**, 2982–2993 (1998).
8. Erdmann, R., Veenhuis, M. & Kunau, W. H. Peroxisomes: organelles at the crossroads. *Trends Cell Biol.* **7**, 400–407 (1997).
9. Hartman, J. J. et al. Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit. *Cell* **93**, 277–287 (1997).
10. Coux, O., Tanaka, K. & Goldberg, A. L. Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* **65**, 801–847 (1996).
11. Larsen, C. N. & Finley, D. Protein translocation channels in the proteasome and other proteases. *Cell* **91**, 431–434 (1997).
12. Lupas, A., Flanagan, J. M., Tamura, T. & Baumeister, W. Self-compartmentalizing proteases. *Trends Biochem. Sci.* **22**, 399–404 (1997).
13. Langer, T. & Neupert, W. Regulated protein degradation in mitochondria. *Experientia* **52**, 1069–1076 (1996).
14. Rep, M. & Grivell, L. A. The role of protein degradation in mitochondrial function and biogenesis. *Curr. Genet.* **30**, 367–380 (1996).
15. Schumann, W. FtsH—a single chain chaperone? *FEMS Microbiol. Rev.* **23**, 1–11 (1999).
16. Casari, G. et al. Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* **93**, 973–983 (1998).
17. Leonhard, K. et al. AAA proteases with catalytic sites on opposite membrane surfaces comprise a proteolytic system for the ATP-dependent degradation of inner membrane proteins in mitochondria. *EMBO J.* **15**, 4218–4229 (1996).
18. Weber, E. R., Hanekamp, T. & Thorsness, P. E. Biochemical and functional analysis of the YME1 gene product, an ATP and zinc-dependent mitochondrial protease from *S. cerevisiae*. *Mol. Biol. Cell* **7**, 307–317 (1996).
19. Pajic, A., Tauer, R., Feldmann, H., Neupert, W. & Langer, T. Yta10p is required for the ATP-dependent degradation of polypeptides in the inner membrane of mitochondria. *FEBS Lett.* **353**, 201–206 (1994).
20. Vestweber, D. & Schatz, G. Point mutations destabilizing a precursor protein enhance its post-translational import into mitochondria. *EMBO J.* **7**, 1147–1151 (1988).
21. Nakai, T., Yasuhara, T., Fujiki, Y. & Ohashi, A. Multiple genes, including a member of the AAA family, are essential for the degradation of unassembled subunit 2 of cytochrome c oxidase in yeast mitochondria. *Mol. Cell Biol.* **15**, 4441–4452 (1995).
22. Pearce, D. A. & Sherman, F. Degradation of cytochrome oxidase subunits in mutants of yeast lacking cytochrome c and suppression of the degradation by mutation of yme1. *J. Biol. Chem.* **270**, 1–4 (1995).
23. Eilers, M. & Schatz, G. Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria. *Nature* **322**, 228–232 (1986).
24. Whiteheart, S. W. et al. N-ethylmaleimide-sensitive fusion protein: a trimeric ATPase whose hydrolysis of ATP is required for membrane fusion. *J. Cell Biol.* **126**, 945–954 (1994).
25. Gottesman, S., Maurizi, M. R. & Wickner, S. Regulatory subunits of energy-dependent proteases. *Cell* **91**, 435–438 (1997).
26. Suzuki, C. K. et al. ATP-dependent proteases that also chaperone protein biogenesis. *Trends Biochem. Sci.* **22**, 118–123 (1997).
27. Wagner, I., Arlt, H., van Dyck, L., Langer, T. & Neupert, W. Molecular chaperones cooperate with PIM1 protease in the degradation of misfolded proteins in mitochondria. *EMBO J.* **13**, 5135–5145 (1994).
28. Huang, S. et al. Effects of conversion of an invariant tryptophan residue to phenylalanine on the function of human dihydrofolate reductase. *Biochemistry* **28**, 471–478 (1989).
29. Tandon, S. & Horowitz, P. M. Reversible folding of rhodanese: presence of intermediate(s) at equilibrium. *J. Biol. Chem.* **264**, 9859–9866 (1989).

Acknowledgements. We thank P. Thorsness for the yme1^{K327R} strain, M. Cahill for the plasmid pGEX-2T-6his-PL2, A. Lupas for sequence alignments and stimulating discussions, and A. Tzagoloff for critically reading the manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft to T. L.

Correspondence and requests for materials should be addressed to T.L. (e-mail: langer@bio.med.uni-muenchen.de).